Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations 1. Thesis and Dissertation Collection, all items

2004-06

Return on investment analysis for the Almond Board of California

Michael, Georgiadis; Drakopoulos, Vassileios; Markopoulos, Thomas

Monterey California. Naval Postgraduate School

http://hdl.handle.net/10945/9927

Downloaded from NPS Archive: Calhoun

Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

MBA PROFESSIONAL REPORT

Return on Investment Analysis for The Almond Board of California

---

By: Drakopoulos Vassileios
Georgiadis Michael
Markopoulos Thomas

June 2004

Advisors: **Kenneth Doerr**
William Gates
Mary Malina

Approved for public release; distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2004	**3. REPORT TYPE AND DATES COVERED** MBA Professional Report

4. TITLE AND SUBTITLE: Return on Investment Analysis for The Almond Board of California

5. FUNDING NUMBERS

6. AUTHOR(S) Drakopoulos Vassileios, Georgiadis Michael, Markopoulos Thomas

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this report are those of the author(s) and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

The purpose of this project is to complete the first phase of a broader research project concerning the exploration of the relationship between consumer attitudes and usage for almonds. It aims to provide the scientific methodology for assessing the effect promotional expenditures have on influencing consumer attitudes as well as relating that to the final impact on the demand for almonds.

This project was conducted with the sponsorship of the Almond Board of California (ABC), an organization supervised by the United States Department of Agriculture, responsible for promoting the consumption and increasing the market share of the California produced almonds in the domestic and international markets.

The primary objective of the project is to develop a methodology that links research, public relations and advertising expenditures made by the ABC, to Attitude, Awareness and Usage (AAU) measurements and eventually to almond shipping and pricing data. The ABC is required by legislation, to conduct a return on investment (ROI) analysis every five years. In conducting this analysis, the organization is interested in developing a management tool that can indicate ROI, but can also be used to identify the portfolio of investments that will maximize AAU (attitudes, awareness, usage). This would allow the ABC to assess the relative impact of its investments portfolio (promotional expenditures) and use this information to make the necessary adjustments in order to improve its effectiveness.

14. SUBJECT TERMS Almond Board of California (ABC), Return on Investment ROI) Analysis, Attitude, Awareness and Usage (AAU), Consumer Attitude

15. NUMBER OF PAGES
121

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

RETURN ON INVESTMENT ANALYSIS FOR THE ALMOND BOARD OF CALIFORNIA

Vassileios Drakopoulos, Lieutenant Commander, Hellenic Navy
Michael Georgiadis, Captain, Hellenic Air Force
Thomas Markopoulos, Major, Hellenic Air Force

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF BUSINESS ADMINISTRATION

from the

NAVAL POSTGRADUATE SCHOOL
June 2004

Authors: ____
Vassileios Drakopoulos, Lieutenant Commander, Hellenic Navy

____
Michael Georgiadis, Captain, Hellenic Air Force

____
Thomas Markopoulos, Major, Hellenic Air Force

Approved by: ____
William Gates, Lead Advisor

____
Kenneth H. Doerr, Support Advisor

____
Mary Malina, Support Advisor

____
Douglas A. Brook, Dean
Graduate School of Business and Public Policy

THIS PAGE INTENTIONALLY LEFT BLANK

RETURN ON INVESTMENT ANALYSIS FOR THE ALMOND BOARD OF CALIFORNIA

ABSTRACT

The purpose of this project is to complete the first phase of a broader research project concerning the exploration of the relationship between consumer attitudes and usage for almonds. It aims to provide the scientific methodology for assessing the effect promotional expenditures have on influencing consumer attitudes as well as relating that to the final impact on the demand for almonds.

This project was conducted with the sponsorship of the Almond Board of California (ABC), an organization supervised by the United States Department of Agriculture, responsible for promoting the consumption and increasing the market share of the California produced almonds in the domestic and international markets.

The primary objective of the project is to develop a methodology that links research, public relations and advertising expenditures made by the ABC, to Attitude, Awareness and Usage (AAU) measurements and eventually to almond shipping and pricing data. The ABC is required by legislation, to conduct a return on investment (ROI) analysis every five years. In conducting this analysis, the organization is interested in developing a management tool that can indicate ROI, but can also be used to identify the portfolio of investments that will maximize AAU (attitudes, awareness, usage). This would allow the ABC to assess the relative impact of its investments portfolio (promotional expenditures) and use this information to make the necessary adjustments in order to improve its effectiveness.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The Almond Board of California (ABC) was established in 1950. It administers a grower-enacted Federal Marketing Order under the supervision of the United States Department of Agriculture. Its role is to promote consumption and to increase the market share of the California produced almonds in the domestic and international markets through generic public relations, advertising, and nutrition research.

An assessment of the marketable kernel pound weight of almonds funds the Board. Existing funds are used to finance key programs including marketing nutrition research, the collection and dissemination of industry statistics, production research, food quality and safety. Thus, the ABC is an organization that invests in certain promotional programs intending to maximize the consumers' interest in almonds, and subsequently, to increase almond consumption. Apparently, the role of the ABC is to use its financial resources effectively to contribute to developing the domestic and international market for almonds.

This is not a clear-cut task, as most of the decisions associated with the allocation of financial resources are made under a great deal of uncertainty because the markets in general function in an extremely complex environment. An infinite number of factors determine their actual position/equilibrium, making the role of the managers even more difficult.

In general, managers rely on the financial theory of supply and demand to make decisions about the market. The value of the theory is well understood and widely accepted. Nevertheless, for organizations such as the ABC, the variety of factors affect supply and demand of which it is very important to be knowledgeable . By understanding the structure of the almond market, managers are able to distinguish between those factors over which they have no control, versus those whose influence can benefit their organization.

Being able to identify any specific factors, as well as isolate their individual effect on the actual demand for almonds, is a very powerful managerial weapon. Even though the idea of totally controlling the market is rather unrealistic, any organization that can manipulate some of the determinants of the market position, such as consumers' attitudes towards almonds, using its financial resources, would like to understand and measure the effects of these manipulations. The resulting benefits are clear. There will be additional profit for the organizations' members and less work for management.

A. DEFINITION OF THE MANAGERIAL PROBLEM-MODEL REQUIREMENTS

The ABC is interested in developing a model to explain the influence of promotional expenditures on consumer attitudes, and consequently, on the usage and demand for almonds. The concept of this model is based on the ABC management's belief that the organization has partial control over specific factors that influence consumer attitudes towards almonds. The term "partial control" refers to the uncertainty involved with respect to the results of using financial resources to affect peoples' attitudes or the effectiveness of promotional expenditures.

At this point, management is interested in determining how a certain expenditure policy affects consumer attitudes and the manner in which these reflect on the final demand for almonds. This conceptual model appears in Figure 1.

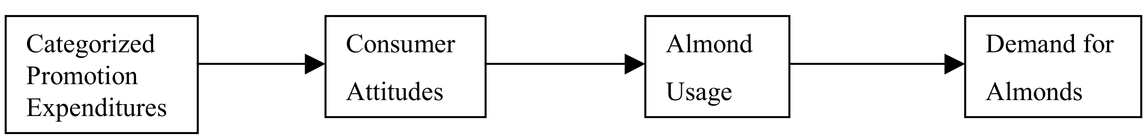


Figure 1. Conceptual Model

The reasoning for this interest is evident. The ABC controls a budget to promote almonds. Consequently, one of the most significant issues for the management of the organization is the optimum allocation of the budgeted funds among the different promotional programs. Keeping in mind that the existence of the organization is based on its ability to promote almonds effectively, the need for a decision making tool that allows for a better assessment of the existing promotional programs is easily understood. A model that identifies the effect a specific promotion program has on the usage/demand for almonds can determine its effectiveness and efficiency. As a result, such a model allows for the optimization of the ABC expenditures portfolio in accordance with the strategic goals established by the management of the organization.

In summary, the scope and the desired capabilities of a model created for the ABC are the following:

- Developing a robust measurement method for the relationship between consumer attitudes constructs and almond usage.
- Mathematically expressing usage as a function of consumer attitudes.
- Measuring the effect that different expenditures categories have on consumer attitudes (effectiveness and efficiency evaluation of the promotional programs).

Such a model will enable the ABC to measure its Return on Investment or the "bang for the buck" for each individual program, to assess its promotional budget and better allocate the organization's financial resources.

B. SCOPE OF THE STUDY

The existing ABC managerial problem was already discussed along with the benefits of resolving it. Nonetheless, the above-presented model is only the idea for a research study, not the actual method leading to the problem's solution. The proper way to proceed with the research is to define a scientific approach to the problem that takes into consideration both the organization's requirements and the existing dataset constraints.

A more detailed examination of the theoretical model reveals some very important issues concerning the needs for specific data that will allow for such an analysis. To develop the described model, the data should be of a similar time interval and have an overlap in terms of their collection periods, which is probably the basic problem associated with this specific research study. Even though the data for promotional expenditures and demand are available and could be easily collected in a timely manner (monthly, quarterly, annually) from the financial statements of the organization, surveys that attempt to measure individual attitudes about almonds, or usage are not conducted as frequently.

To overcome the inconsistency problem between the datasets, it is necessary to modify the conceptual model and examine the issues described above in two different phases:

- Phase a. The relationship between attitudes and usage, which will be based on the existing datasets from the conducted surveys and
- Phase b. The relationship of categories of expenditures and actual demand of almonds, which will be based on the organization's financial data and shipments/prices of almonds.

According to this proposal, the modified conceptual model appears in Figure 2.

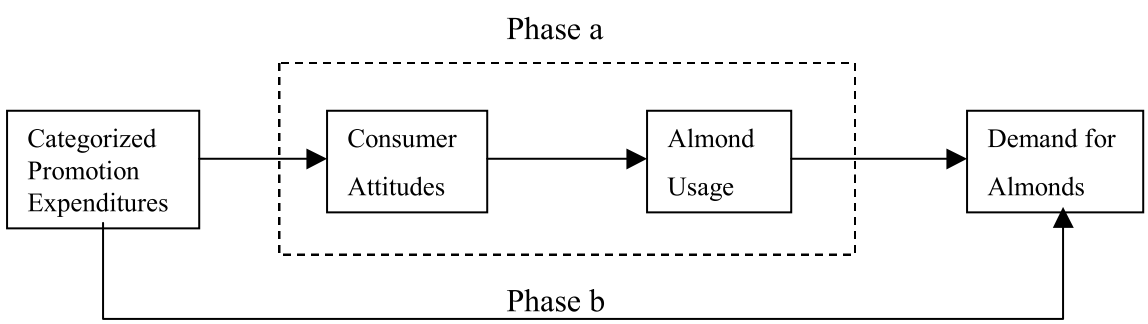


Figure 2. Conceptual Model's Phases

The purpose of this project is to complete the first research phase, which is the exploration of the relationship between consumers' attitudes and usage for almonds and to provide the scientific methodology for processing the second phase. Although it seems to be a straightforward analysis, in reality there are certain necessary issues to address before using mathematical or statistical tools to develop the actual model.

There are indeed some difficulties related to the early stages of this research. First of all, the provided questionnaires that will be used to extract data for the analysis were not developed for the specific purposes of this research. As a result, the datasets created by the answers to the questions include both quantitative and qualitative variables in a variety of scales that either prohibit their use or require further transformations before using them in a mathematical/statistical model. Secondly, attitudes cannot be measured directly. Thus, a part of the analysis should be the development of constructs that will allow for the proper measurement of these attitudes.

Overcoming these difficulties will allow for the best possible use of the provided datasets towards reaching the following goals:

- Identification of common/underlying factors from the survey questionnaires that could provide useful information about consumer's attitudes toward almonds,
- Development of a methodology for the robust measurement of the constructs created to describe consumers' attitudes, and
- Identification and mathematical representation of the relationship between potentially existing factors explaining consumer attitudes and usage of almonds.

C. METHODOLOGY

In analyzing the model for the ABC, it is necessary to follow a sequence in the research processes in order to approach the problem described above and eventually to identify all critical parameters that directly affect consumer attitudes, and consequently, the behavior of the almond market. The intermediate stage of the model requires two abstract concepts as components that cannot be presented using mathematical methods. This suggests a difficulty in developing a model that will describe the situation. In order to overcome this difficulty, the statistical process of factor analysis is used as an appropriate research method. The first step of this analysis is to explore the market context through the individual questionnaire responses and to identify any implied factors relative to consumer attitudes. Implementing factor analysis will identify certain latent variables and parameters that are components of the hypothesized model attempting to describe the relationship between two or more abstract entities, such as consumer attitudes and almond usage.

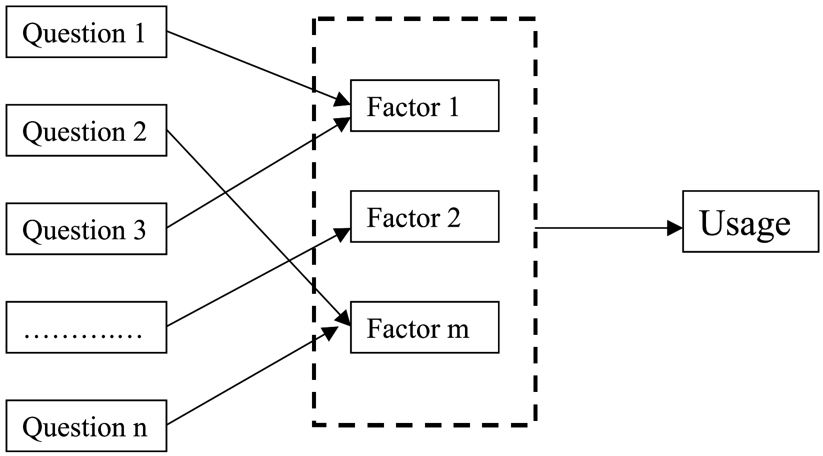


Figure 3. Relationship between Two or More Abstract Entities

The most important contribution of this process is grouping questions with similar meaning that identify a specific factor, expected to be the implied driver of consumer behavior. Therefore, the use of exploratory factor analysis will allow the transformation of non-tangible and abstract concepts, such as human attitudes, into quantifiable constructs easily usable for further statistical analysis.

After defining the factor model, the second step of the analysis is to assess the relationship between the constructed factor scores and almond usage. A proper statistical method is regression analysis. According to that process, the individual factor scores are entered into a linear model as independent variables presumed to be the predictors of the dependent variable of almond usage. Initially, the analysis evaluates whether a statistically significant relationship exists between the variables. If that is established, then a linear regression equation is created representing mathematically the estimated relationship. The major benefit is that the usage variable can be expressed as a function of the factor score coefficients. Once the weight of each one of those factor score coefficients is determined, the question of how individual attitudes affect the usage of almonds is answered.

The second phase of the research uses the same linear approach attempting to explain the relationship between the different promotional expenditure categories and the demand for almonds. In this case, the goal is to develop a methodology that will allow the ABC management to use the existing financial data efficiently to measure the effectiveness of its promotional programs.

Understanding that the linear model described above along with the single financial constraint of the budget ceiling cannot provide a unique solution to a linear programming problem searching for the optimum portfolio of expenditures, it is possible to conclude that the model will only be useful to identify the relative impact of each expenditure category on the demand for almonds. Nevertheless, the developed equation/model (if validated) will still provide a proper method to address the problem of assessing the promotional program effectiveness, simple and number oriented, making it an extremely powerful decision-making tool for the ABC management.

THIS PAGE INTENTIONALLY LEFT BLANK

II. EXPLORATORY FACTOR ANALYSIS

A. BUSINESS QUESTION

The introduction presented the conceptual model of the research. This chapter examines the relationship between consumers' attitudes to almonds and their usage of almonds. For the Almond Board of California, isolating and measuring the effect that a certain attitude has on the usage of almonds is a valuable financial tool. Indeed, by understanding how a specific attitude could influence the usage/consumption of almonds, the ABC will be able to manage its expenditures/focus its investments towards certain areas better.

To examine the relationship described above, the ABC provided three datasets consisting of the individual answers to the three AAU (Attitude, Awareness, Usage) surveys conducted in 1999, 2001 and 2003. Even though the design of the questionnaires reflected marketing purposes, rather than the sort of behavioral study undertaken in this thesis, the main idea is to use those studies efficiently in order to extract the best possible information.

Based on the belief that certain answers to the AAU questionnaires could uncover real attitude dimensions, exploratory factor analysis will be used to identify and team those sets of questions that identify a specific direction. All questions and their respective answers that could provide meaningful data for analysis will be introduced to the method as variables in an attempt to extract a set of factors that represents underlying but not explicitly visible consumer attitudes.

The goals of the implementation of the described method above are:

- Examine whether the AAU surveys-questionnaires could be used to identify and provide measurements for factors that potentially influence the usage of almonds,
- Develop a model for measuring these factors as well as assessing the quality of the developed measurements,
- Examine whether a statistically significant relationship exists between the extracted factors and usage and if it does, and
- Assess the validity of the model and its forecasting ability.

Achieving these goals will provide a clear view of the attitudes-usage interaction mechanism that will allow for better decisions concerning budgeting for promotion programs and allocation of the existing financial resources among them.

B. NEED FOR MEASUREMENTS

There are two main problems associated with the development of a model that aims to describe the nature of the relationship between consumers' attitudes and the usage of almonds. The first is the fact that no clear rules related to the identification/categorization of a specific attitude exists, and the second is that attitudes cannot be easily quantified and measured. Practically, most measurement is indirect, meaning that the phenomenon of the interest is not measured directly but is instead inferred from specific indicators.

In this case, the questions from the three AAU surveys are the only available data that could be potentially used as indicators of attitudes. As a result, in order to be able to process the existing data and extract useful information, it is necessary to agree primarily on a way to convert or recode them into a meaningful scale.

The results of this filtering or variable synthesizing process will allow for mapping the existing set of constructs onto a set of numbers that will be valuable for measurement, interpretation, presentation and possibly forecasting purposes.

C. DATA FILTERING

The original datasets consisted of three SPSS files containing the individual answers to the three AAU surveys. Slight changes and adjustments made to the surveys from year to year result in differences among the three questionnaires. In order to have a common basis (comparability in results) for the analysis, the authors conducted an initial screening of the datasets isolating common questions and adjusting individual answers to meaningful scales with respect to almonds.

The sample sizes were 750, 753 and 700 for the years 1999, 2001 and 2003 respectively. From the aforementioned datasets, it was possible to extract 32 different questions and individual responses that could provide meaningful data for further processing. Questions irrelevant to the purposes of this study were omitted, such as screening questions or questions referring to the characteristics of the individuals in the sample. To improve comparability across surveys, the authors also excluded questions from the early datasets discontinued in later surveys and not replaced with questions similar in meaning.

Appendix A presents the selected questions and the transformations required for processing. The set of questions in Appendix A represent all-available extracted data that could be potentially used to evaluate or assess attitudes towards almonds. Its actual usability or usefulness has not yet been determined, as this is only an attempt to create a functional dataset for statistical research and analysis. Some of the questions will be used individually as variables in the analysis and others will be combined to create measurements for other developed constructs.

As one of the purposes of the research is to examine the usage of almonds, it is necessary to identify a variable that will be used to measure this construct. In the provided datasets, only two questions were found to be relevant, both referring to the frequency of almond purchases rather than actual quantities purchased:

- "About how many times per MONTH do you purchase almonds, in any type or form?" and
- [If less than once a month] "About how often do you purchase almonds?" with the following answer options:
    - About once every 2-3 months
    - Once every 4-6 months
    - Only once every 7 months to one year
    - Less than once a year
    - Never
    - DK/NS (Don't know/not sure)

Using these two questions, it was possible to create the following two variables (uf, unf) related to usage.

- uf: Stands for usage measured in non-integer numbers, including fractions. (Usage with Fractions). Purchases for less than once a month are included. The variable was created by combining answers from the above questions. For non-frequent purchasers, (individuals that answered zero to question 1 above) a fraction is used to represent the monthly almond purchases based on their answers to question two above, or zero for those that answered zero or never to both questions. DK/NS answers were handled as missing values.
- unf: Stands for usage measured in integer numbers, excluding fractions. (Usage non-Fractioned). Purchases for less than once a month are excluded. The variable takes into consideration only frequent purchasers. It includes the answers to question 1 above, handling the non-frequent purchasers (those who answered less than one a month in question 1), DK/NS and missing answers as missing values.

An explanation of the recoding process from the raw datasets appears below:

- Step 1: All answers other than 0, DK/NS, missing, were used as given. (Both for uf, unf variables)
- Step 2: (only for the uf variable) The answer to question 2 above was recoded as follows:
    - About once every 2-3 months: 1/ 2.5 =0.4
    - Once every 4-6 months: 1/ 5 =0.2
    - Only once every 7 months to one year: 1/ 9.5=0.105
    - Less than once a year: 1/ 18=0.0556
    - Never:0
    - DK/NS: missing values

The following table summarizes the results of the clearing or filtering process of the three datasets, the measurement scales of the data and their use:

Year	0-1 Scale	Ratio Scale
1999	1,3,4,6,7,8,9,10,11,12,13,14,15,16	2,17,18,19,20,21,22,23,25,27,28,29,30,31
2001	1,3,4,6,7,8,9,10,11,12,13,14,15,16	2,17,18,19,20,21,22,23,24,26,27,28,29,30,31
2003	1,3,4,5,6,7,8,9,10,11,12,13, 14,15,16	2,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32
Common Questions		
	1,3,4,6,7,8,9,10,11,12,13,14,15,16	2,17,18,19,20,21,22,23,27,28,29,30,31

Table 1. Questions Identified in Each Dataset

The above table presents the 0-1 scale questions separately, as it is only be possible to use them for the synthesis of other variables while the ratio scale questions individually as variables in factor analysis.

D. ANALYSES

A general approach for the analysis is first to identify relevant factors concerning consumer behavior using exploratory factor analysis (EFA) and then to insert the identified factors as independent variables in a regression equation in order to evaluate their relationship with usage. A scientifically valid procedure to assess the quality of the results from the factor analysis (validate the extracted model) before using them any further is to conduct a reliability test (Cronbach's alpha) for the developed factor model from the 1999 dataset using the other two datasets (2001, 2003). The results of this test will potentially validate the quality of the developed measurement scales, their internal consistency, and the extent to which the selected questions/variables are related to each other.

1. Identifying Factors Related to Attitudes

In an attempt to identify potential sets of questions that could explain underlying dimensions or attitudes towards almonds and provide relevant measurements, the authors first examined the 1999 dataset using exploratory factor analysis. Initially all the common (ratio scale) questions among the three datasets (see Table in paragraph C above) were entered into the analysis. The SPSS settings, the extraction and rotation methods used, the calculation of factors scores and the used criteria for determining and assessing the results were the same in all analyses. A summary follows.

a. Factor Extraction Method

The Principal Axis Factoring method was used.[1] The specific method was selected as one of the preferred techniques for factor analysis. Although the method restricts conclusions to the sample collected, the idea is to see if the analysis can provide the same results concerning the factors' structure from the three different survey datasets, indicating that it was possible to generate the conclusions to the consumers' population.

[1] Pedhazur and Schmelkin, 598.

The factor extraction was to be based on the analysis of the correlation matrix that compensates for the use of variables measured in different scales (1-10 scale vs. 1-5 scale). The option selected pertaining to the retention of factors was that of Eigenvalues over 1.[2] Also, a scree plot was requested as a second criterion for determining the appropriate number of factors to retain in the model.

b. Rotation

In order to improve the interpretability of factors, the authors selected the Varimax rotation method. Varimax is an orthogonal rotation that attempts to maximize the dispersion of loadings within factors. This actually is an attempt to load a smaller number of variables highly onto each factor resulting in a more interpretable or easy method to identify clusters of factors.[3]

c. Factor Scores

For the calculation of factor scores, the authors selected the Anderson-Rubin method. This method produces factor scores that are uncorrelated and standardized with a mean of 0 and a standard deviation of 1. The basic reason for this choice is the potential use of factor scores as independent variables in a multiple linear regression analysis that attempts to predict usage and, in that sense, could assist in the implementation of the model and improve the objectivity of the results.[4]

d. Other Options

For the missing values, the authors selected the "Exclude cases listwise" option mainly because of the large sample size and the randomness of missing data. For the display of coefficients in outputs, the authors selected the "Suppress absolute values less than" option, setting the value to .40 to facilitate interpretation.

e. Screening

The correlation matrices in Tables 2, 3, and 4 were produced from the first attempt. These tables were produced using all the common questions and the usage variables unf and uf respectively. The top half of each table contains the Pearson correlation coefficient between all pairs of questions included in the analysis, whereas the

[2] Kaiser's recommendation.

[3] Kim, Jae-on, and Mueller, C., 34.

[4] Field, 431.

bottom half contains the one-tailed significance of these coefficients. Based on the idea that it is necessary have variables that correlate fairly well to do a factor analysis, it is possible to use the matrices to identify variables that do not correlate with others and should be excluded from the model.

The screening of the variables was conducted considering the correlation coefficient matrix and their corresponding significance levels. The majority of the correlation coefficients for any variable were low and the majority of the significance values were greater than 0.05.

Scanning the significance values, it was concluded that questions 20, 27, unf and uf do not correlate well with the other variables and should be considered for elimination from the model. The practical reason for excluding the aforementioned variables lies in the way the questions are stated. Specifically, for question 20, (*Thinking specifically about the healthfulness of almonds, on a 1 to 10 scale, where 1 means you don't think almonds are healthy at all and 10 means you think almonds are very healthy, which number best describes your opinion?*) note that it is more generic than other questions in respect to nutritional perceptions increasing the probability of ambiguous responses. On the contrary, question 27 (*Thinking only about cholesterol and almonds, do you think almonds are high, moderate or low in cholesterol, or do almonds contain no cholesterol at all? If you aren't sure, just say so),* requires specialized medical-nutrition knowledge about almonds and their effects on human health that is not common to the majority of the individuals in the sample, resulting in a higher probability of wrong answers.

The reason that the unf and uf variables, (*And about how many times per MONTH do you purchase almonds, in any type or form?*) are not well correlated with the other variables is that they are probably the only questions that refer to usage in general. These results were anticipated because the other questions express perceptions about almonds whereas the unf and uf variables refer to the effect of those perceptions.

By eliminating the above-mentioned questions (Q20 and Q27), the need for highly correlated questions before running factor analysis were met. Table 3 presents the correlation matrix used as input to the factor analysis. The output of the factor analysis appears next.

Correlation Matrix

Correlation

Q	2	17	18	19	20	21	22	23	27	28	29	30	31	UNF
2	1	0.418	0.416	0.4	-0.055	0.193	0.215	0.155	0.052	0.16	0.077	0.058	0.099	-0.019
17	0.418	1	0.922	0.726	-0.09	0.47	0.406	0.28	0.056	0.212	0.11	0.152	0.156	0.007
18	0.416	0.922	1	0.734	-0.071	0.481	0.4	0.286	0.042	0.219	0.13	0.143	0.148	0.012
19	0.4	0.726	0.734	1	-0.031	0.426	0.346	0.331	0.082	0.236	0.099	0.121	0.184	0.032
20	-0.055	-0.09	-0.071	-0.031	1	0.006	-0.094	0.066	-0.035	-0.036	-0.087	-0.083	-0.101	0.056
21	0.193	0.47	0.481	0.426	0.006	1	0.496	0.395	-0.011	0.267	0.085	0.118	0.192	0.07
22	0.215	0.406	0.4	0.346	-0.094	0.496	1	0.368	0.026	0.195	0.122	0.094	0.064	0.013
23	0.155	0.28	0.286	0.331	0.066	0.395	0.368	1	-0.023	0.166	0.018	0.058	0.051	0.051
27	0.052	0.056	0.042	0.082	-0.035	-0.011	0.026	-0.023	1	0.179	0.308	0.142	0.059	0.036
28	0.16	0.212	0.219	0.236	-0.036	0.267	0.195	0.166	0.179	1	0.318	0.28	0.394	-0.026
29	0.077	0.11	0.13	0.099	-0.087	0.085	0.122	0.018	0.308	0.318	1	0.216	0.148	-0.085
30	0.058	0.152	0.143	0.121	-0.083	0.118	0.094	0.058	0.142	0.28	0.216	1	0.236	0.047
31	0.099	0.156	0.148	0.184	-0.101	0.192	0.064	0.051	0.059	0.394	0.148	0.236	1	0.02
UNF	-0.019	0.007	0.012	0.032	0.056	0.07	0.013	0.051	0.036	-0.026	-0.085	0.047	0.02	1

Sig. (1-tailed)

Q	2	17	18	19	20	21	22	23	27	28	29	30	31	UNF
2		0.000	0.000	0.000	0.107	0.000	0.000	0.000	0.122	0.000	0.041	0.097	0.013	0.335
17	0.000		0.000	0.000	0.021	0.000	0.000	0.000	0.102	0.000	0.006	0.000	0.000	0.435
18	0.000	0.000		0.000	0.055	0.000	0.000	0.000	0.173	0.000	0.002	0.001	0.000	0.391
19	0.000	0.000	0.000		0.240	0.000	0.000	0.000	0.031	0.000	0.013	0.003	0.000	0.238
20	0.107	0.021	0.055	0.240		0.446	0.017	0.069	0.218	0.208	0.024	0.031	0.011	0.102
21	0.000	0.000	0.000	0.000	0.446		0.000	0.000	0.404	0.000	0.028	0.004	0.000	0.057
22	0.000	0.000	0.000	0.000	0.017	0.000		0.000	0.275	0.000	0.003	0.017	0.076	0.388
23	0.000	0.000	0.000	0.000	0.069	0.000	0.000		0.306	0.000	0.343	0.095	0.124	0.125
27	0.122	0.102	0.173	0.031	0.218	0.404	0.275	0.306		0.000	0.000	0.001	0.091	0.206
28	0.000	0.000	0.000	0.000	0.208	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.277
29	0.041	0.006	0.002	0.013	0.024	0.028	0.003	0.343	0.000	0.000		0.000	0.000	0.028
30	0.097	0.000	0.001	0.003	0.031	0.004	0.017	0.095	0.001	0.000	0.000		0.000	0.144
31	0.013	0.000	0.000	0.000	0.011	0.000	0.076	0.124	0.091	0.000	0.000	0.000		0.330
UNF	0.335	0.435	0.391	0.238	0.102	0.057	0.388	0.125	0.206	0.277	0.028	0.144	0.330	

a Determinant = 1.058E-02

Table 2. Correlation Matrix with UNF Variable.

Correlation Matrix

Correlation

Q	2	17	18	19	20	21	22	23	27	28	29	30	31	UF
2	1	0.459	0.451	0.413	-0.036	0.219	0.205	0.188	0.048	0.183	0.081	0.05	0.093	-0.022
17	0.459	1	0.927	0.726	-0.047	0.467	0.397	0.301	0.048	0.218	0.087	0.122	0.146	0.011
18	0.451	0.927	1	0.727	-0.031	0.485	0.395	0.296	0.041	0.229	0.107	0.114	0.152	0.008
19	0.413	0.726	0.727	1	-0.013	0.44	0.389	0.344	0.055	0.225	0.073	0.113	0.186	0.017
20	-0.036	-0.047	-0.031	-0.013	1	-0.008	-0.079	0.039	-0.028	-0.047	-0.074	-0.04	-0.107	0.046
21	0.219	0.467	0.485	0.44	-0.008	1	0.56	0.436	0.024	0.253	0.059	0.094	0.195	0.019
22	0.205	0.397	0.395	0.389	-0.079	0.56	1	0.406	0.041	0.208	0.07	0.095	0.13	-0.012
23	0.188	0.301	0.296	0.344	0.039	0.436	0.406	1	-0.036	0.2	-0.022	0.079	0.106	0.011
27	0.048	0.048	0.041	0.055	-0.028	0.024	0.041	-0.036	1	0.186	0.367	0.11	0.033	0.02
28	0.183	0.218	0.229	0.225	-0.047	0.253	0.208	0.2	0.186	1	0.294	0.29	0.409	-0.028
29	0.081	0.087	0.107	0.073	-0.074	0.059	0.07	-0.022	0.367	0.294	1	0.169	0.123	-0.069
30	0.05	0.122	0.114	0.113	-0.04	0.094	0.095	0.079	0.11	0.29	0.169	1	0.257	0.046
31	0.093	0.146	0.152	0.186	-0.107	0.195	0.13	0.106	0.033	0.409	0.123	0.257	1	-0.009
UF	-0.022	0.011	0.008	0.017	0.046	0.019	-0.012	0.011	0.02	-0.028	-0.069	0.046	-0.009	1

Sig. (1-tailed)

Q	2	17	18	19	20	21	22	23	27	28	29	30	31	UF
2		0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.099	0.000	0.015	0.091	0.007	0.275
17	0.000		0.000	0.000	0.105	0.000	0.000	0.000	0.099	0.000	0.010	0.001	0.000	0.382
18	0.000	0.000		0.000	0.206	0.000	0.000	0.000	0.138	0.000	0.002	0.001	0.000	0.418
19	0.000	0.000	0.000		0.366	0.000	0.000	0.000	0.071	0.000	0.026	0.001	0.000	0.325
20	0.167	0.105	0.206	0.366		0.411	0.017	0.150	0.227	0.103	0.023	0.140	0.002	0.111
21	0.000	0.000	0.000	0.000	0.411		0.000	0.000	0.261	0.000	0.057	0.006	0.000	0.305
22	0.000	0.000	0.000	0.000	0.017	0.000		0.000	0.136	0.000	0.031	0.006	0.000	0.378
23	0.000	0.000	0.000	0.000	0.150	0.000	0.000		0.166	0.000	0.277	0.018	0.002	0.386
27	0.099	0.099	0.138	0.071	0.227	0.261	0.136	0.166		0.000	0.000	0.002	0.191	0.298
28	0.000	0.000	0.000	0.000	0.103	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.229
29	0.015	0.010	0.002	0.026	0.023	0.057	0.031	0.277	0.000	0.000		0.000	0.000	0.033
30	0.091	0.001	0.001	0.001	0.140	0.006	0.006	0.018	0.002	0.000	0.000		0.000	0.110
31	0.007	0.000	0.000	0.000	0.002	0.000	0.000	0.002	0.191	0.000	0.000	0.000		0.405
UF	0.275	0.382	0.418	0.325	0.111	0.305	0.378	0.386	0.298	0.229	0.033	0.110	0.405	

a Determinant = 8.591E-03

Table 3. Input to the Factor Analysis.

Correlation Matrix

Correlation

Q	2	17	18	19	21	22	23	28	29	30	31
2	1	0.464	0.455	0.423	0.227	0.215	0.193	0.178	0.074	0.046	0.082
17	0.464	1	0.929	0.728	0.475	0.400	0.305	0.21	0.086	0.118	0.142
18	0.455	0.929	1	0.730	0.492	0.397	0.302	0.219	0.105	0.109	0.147
19	0.423	0.728	0.730	1	0.448	0.397	0.350	0.214	0.059	0.108	0.173
21	0.227	0.475	0.492	0.448	1	0.562	0.439	0.251	0.056	0.092	0.189
22	0.215	0.400	0.397	0.397	0.562	1	0.409	0.208	0.061	0.093	0.126
23	0.193	0.305	0.302	0.350	0.439	0.409	1	0.201	-0.027	0.076	0.102
28	0.178	0.210	0.219	0.214	0.251	0.208	0.201	1	0.285	0.287	0.405
29	0.074	0.086	0.105	0.059	0.056	0.061	-0.027	0.285	1	0.165	0.123
30	0.046	0.118	0.109	0.108	0.092	0.093	0.076	0.287	0.165	1	0.259
31	0.082	0.142	0.147	0.173	0.189	0.126	0.102	0.405	0.123	0.259	1

Sig. (1-tailed)											
Q	**2**	**17**	**18**	**19**	**21**	**22**	**23**	**28**	**29**	**30**	**31**
2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.107	0.014
17	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.001	0.000
18	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.000
19	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.055	0.002	0.000
21	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.064	0.006	0.000
22	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.050	0.006	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.234	0.019	0.003
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
29	0.023	0.010	0.002	0.055	0.064	0.050	0.234	0.000		0.000	0.000
30	0.107	0.001	0.002	0.002	0.006	0.006	0.019	0.000	0.000		0.000
31	0.014	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	
a	Determinant = 1.013E-02										

Table 4. Correlation Matrix without UF and UNF Variable.

2. Factor Analysis Results

As seen in Tables 2-4, all correlation coefficients are relatively high and the majority of the significance values for any variable are low (less than 0.05), indicating that the selected variables correlate fairly well but not perfectly with all others and are appropriate for factor analysis. The determinant at the bottom of the matrix is 0.01013 greater than the necessary value of 0.00001 showing that no singularity problem with the data exists and multicollinearity is not a problem.

Having already seen that the selected variables are appropriate for factor analysis, the next consideration is whether the sample size is adequate. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is an indicator of sample adequacy.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.814
Bartlett's Test of Sphericity	Approx. Chi-Square	3340.631
	df	55
	Sig.	.000

Table 5. KMO and Bartlett's Test.

The value of the KMO statistic is 0.814 as shown in Table 5, which is relatively close to 1.0 and indicates that patterns of correlations are relatively compact. Therefore, factor analysis is expected to yield distinct and reliable factors. Actually, according to Kaiser (1974) values between 0.8 and 0.9 are considered “great” for analysis.[5] Therefore, it is possible to be confident that factor analysis is appropriate for these data.

Bartlett’s measure of Sphericity also confirms that the sample is appropriate for factor analysis (χ^2 = 3340.6, df = 55, p < 0.000) and consequently, that are some relationships between variables that will be included in the analysis.

The next step is the identification of the number of factors that should be included in the model. Having selected eleven questions for the model, it is known that there are also eleven factors in the R matrix that explain 100% of the variance. Nevertheless, most are expected to be unimportant. The criterion used to determine the importance of each factor, and consequently, to assess which factors to retain and to discard is based on the eigenvalues. As already noted, Kaiser (1970) recommends retaining factors with eigenvalues greater than 1.0. Table 6 presents the eigenvalues associated with each factor before extraction, after extraction and after rotation.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.001	36.376	36.376	3.665	33.318	33.318	2.582	23.472	23.472
2	1.570	14.276	50.651	.982	8.928	42.246	1.547	14.064	37.535
3	1.142	10.377	61.029	.654	5.944	48.190	1.172	10.654	48.190
4	.878	7.986	69.015						
5	.771	7.012	76.027						
6	.707	6.427	82.454						
7	.588	5.342	87.795						
8	.526	4.786	92.581						
9	.425	3.864	96.445						
10	.321	2.920	99.365						
11	6.984E-02	.635	100.000						

Extraction Method: Principal Axis Factoring.

Table 6. Factor Eigenvalues.

Note that by using the Kaiser’s criterion, SPSS has already extracted the first three factors with eigenvalues greater than 1.0, explaining 48.190% of total variance. The extracted factors explain 33.318%, 8.928% and 5.944% of the variance respectively.

[5] Field, 455.

After rotation, the factor structure is optimized and the relative importance of factors is better balanced. Factor 1 then explains 23.472% of the variance, compared to factors 2 and 3 that explain 14.064% and 10.654%, respectively.

Nevertheless, Kaiser's criterion used for the extraction of the three factors is accurate when there are less than 30 variables and communalities after extraction are greater than 0.7 or when the sample size exceeds 250 and the average communality is greater than 0.6. In this case, (sample size >250), as shown in Table 6, there are some communalities that exceed 0.7 but the average remains below 0.6, indicating that Kaiser's criterion for extracting factors may not be accurate.[6]

The last method to be considered for estimating the number of factors to be extracted is the evaluation of the scree plot. According to this criterion/method, the number of factors to be retained should be equal to the point of inflection on the curve. Even though it is not possible to identify the exact point clearly in Figure 1, it is probable to justify retaining three or four factors. Finally, given the size of the sample and the results of both criteria, the conclusion is that retaining three factors is most likely safe.

Communalities		
Question	**Initial**	**Extraction**
2	0.241	0.246
17	0.871	0.922
18	0.873	0.916
19	0.581	0.613
21	0.446	0.597
22	0.371	0.505
23	0.263	0.341
28	0.29	0.576
29	0.103	0.129
30	0.119	0.18
31	0.198	0.277

Extraction Method: Principal Axis Factoring.

Table 7. Communalities

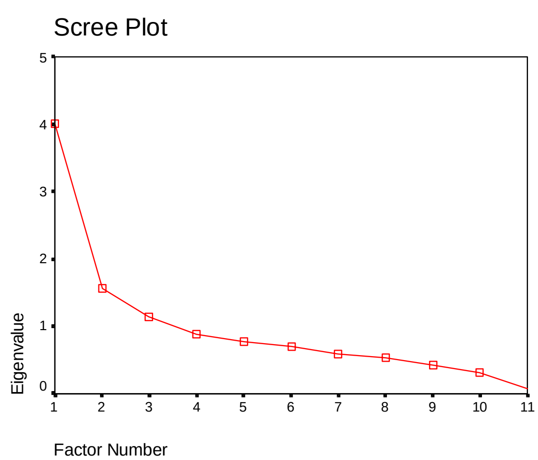


Figure 4. Scree Plot

To assess the goodness of fit of the constructed model (3 factors retained), the reproduced correlation matrix (Table 8) were examined. The top half of this matrix

[6] Field, p. 457.

contains the correlation coefficients between all the questions based on the factor model. If the model were a perfect fit of the data, it is expected that the reproduced correlation coefficients would be the same as the original correlation coefficients. The lower half of the reproduced matrix contains the differences between the observed correlation coefficients and those predicted from the model. As a result, it is desired that most of these values be small (<0.05).

To facilitate the evaluation of the residual matrix, SPSS provides a footnote summary that states the number of residuals that have an absolute value greater than the threshold value (0.05). Indeed, only one residual (1% of total residuals) is greater than 0.05 indicating that there are no concerns about the goodness of fit of the model.

Reproduced Correlations											
	Reproduced Correlation										
Q	**2**	**17**	**18**	**19**	**21**	**22**	**23**	**28**	**29**	**30**	**31**
2	0.246	0.473	0.472	0.383	0.255	0.216	0.171	0.151	0.064	0.076	0.102
17	0.473	0.922	0.919	0.737	0.469	0.396	0.314	0.214	0.090	0.105	0.143
18	0.472	0.919	0.916	0.736	0.473	0.400	0.317	0.222	0.094	0.109	0.149
19	0.383	0.737	0.736	0.613	0.466	0.404	0.324	0.223	0.079	0.105	0.150
21	0.255	0.469	0.473	0.466	0.597	0.547	0.449	0.258	0.044	0.105	0.173
22	0.216	0.396	0.400	0.404	0.547	0.505	0.415	0.217	0.028	0.084	0.144
23	0.171	0.314	0.317	0.324	0.449	0.415	0.341	0.175	0.020	0.067	0.116
28	0.151	0.214	0.222	0.223	0.258	0.217	0.175	0.576	0.256	0.319	0.399
29	0.064	0.090	0.094	0.079	0.044	0.028	0.020	0.256	0.129	0.149	0.179
30	0.076	0.105	0.109	0.105	0.105	0.084	0.067	0.319	0.149	0.180	0.221
31	0.102	0.143	0.149	0.150	0.173	0.144	0.116	0.399	0.179	0.221	0.277
	Residual										
2		-0.009	-0.017	0.040	-0.028	-0.001	0.022	0.027	0.010	-0.030	-0.021
17	-0.009		0.011	-0.010	0.006	0.004	-0.009	-0.004	-0.004	0.013	-0.001
18	-0.017	0.011		-0.007	0.019	-0.002	-0.016	-0.003	0.011	0.000	-0.002
19	0.040	-0.010	-0.007		-0.018	-0.008	0.026	-0.009	-0.020	0.003	0.023
21	-0.028	0.006	0.019	-0.018		0.014	-0.010	-0.007	0.012	-0.013	0.016
22	-0.001	0.004	-0.002	-0.008	0.014		-0.006	-0.009	0.033	0.009	-0.018
23	0.022	-0.009	-0.016	0.026	-0.010	-0.006		0.026	-0.047	0.010	-0.014
28	0.027	-0.004	-0.003	-0.009	-0.007	-0.009	0.026		0.028	-0.032	0.005
29	0.010	-0.004	0.011	-0.020	0.012	0.033	-0.047	0.028		0.016	-0.055
30	-0.030	0.013	0.000	0.003	-0.013	0.009	0.010	-0.032	0.016		0.038
31	-0.021	-0.001	-0.002	0.023	0.016	-0.018	-0.014	0.005	-0.055	0.038	

Extraction Method: Principal Axis Factoring.

a Residuals are computed between observed and reproduced correlations. There are 1 (1.0%) nonredundant residuals with absolute values > 0.05.

b Reproduced communalities

Table 8. Correlation Residuals.

Tables 9 and 10 present the unrotated and rotated factor loadings for each variable. With the exception of variable/question # 29, factor loadings below 0.4 are not displayed for the rotated factor matrix. The criterion behind suppressing loadings less than 0.4 is based on Stevens's (1992) suggestion that this cut-off point is appropriate for interpretative purposes.

Unrotated Factor Matrix			
	Factor		
Question	**1**	**2**	**3**
2	**0.473**		0.123
17	**0.892**	-0.266	0.236
18	**0.894**	-0.252	0.231
19	**0.768**	-0.136	
21	**0.657**		-0.397
22	**0.575**		-0.41
23	**0.464**		-0.348
28	0.392	**0.635**	0.14
29	0.138	0.28	0.178
30	0.189	0.357	0.129
31	0.265	**0.442**	0.103

Extraction Method: Principal Axis Factoring.
a 3 factors extracted. 16 iterations required.

Table 9. Unrotated Factor Matrix

Rotated Factor Matrix			
	Factor		
Question	**1**	**2**	**3**
2	0.462		
17	0.921		
18	0.914		
19	0.693		
21		0.698	
22		0.662	
23		0.549	
28			0.729
29			**0.352***
30			0.418
31			0.507

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 5 iterations.

Table 10. Rotated Factor Matrix

Comparing the results from the rotated factor matrix with the unrotated solution, notice that before rotation most variables are loaded highly on the first factor. However, after rotation, the factor structure has clarified things considerably. Now, the questions have formed three different groups, each one loading highly onto one single factor. Questions 2,17,18,19 load highly onto the first factor, questions 21, 22, 23 load onto the second and questions 28,29,30,31 load onto the third factor respectively.

Looking at the content of the questions that load onto the same factor, it is possible to identify common themes. In fact, questions that load highly on factor #1 seem to all relate to a general liking for almonds, while those that load highly on factors #2 and #3 seem to all relate to general knowledge and specific nutritional perceptions about almonds respectively. To facilitate the reporting and interpretation of the results and to maintain consistency in this research, henceforth, #1 factor will be labeled as "liking", #2 as "awareness" and #3 as "nutritional perceptions".

Finally, to assess the appropriateness of the selected orthogonal rotation, it is necessary to the factor transformation matrix presented below:

Factor Transformation Matrix			
Factor	**1**	**2**	**3**
1	0.787	0.543	0.293
2	-0.432	0.146	0.89
3	0.44	-0.827	0.349

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

Table 11. Factor Transformation Matrix

The matrix indicates that some rotation was necessary to obtain a sufficient solution. Otherwise, this matrix would be an identity matrix. A symmetrical matrix (same values above and below the diagonal) would be expected if orthogonal rotation were completely appropriate. Even though the values above and below the diagonal are not identical, the matrix cannot be considered as "very" unsymmetrical providing a reason to try oblique rotation (assumption of correlated factors). Consequently, the application of orthogonal rotation can be considered satisfactory.[7]

Having concluded that a meaningful measurement model was developed, examining the data from 1999, the next step is to assess the reliability of the model using the available 2001 and 2003 data. Using the same assumptions and settings as for 1999, the authors conducted factor analysis to the datasets of 2001 and 2003. Appendix B presents the output of this analysis.

All the previous analyses converge at the same groups of questions for determining the consumers' attitudes. Therefore, a safe method to measure the identified attitudes is to use the factor scores derived from the individual responses to the survey questions. According to the results of the research, the sets of questions provided in Appendix C correspond to each factor and should be used for measurement over time.

3. Cronbach's Alpha – Reliability Test

The developed datasets from the questionnaires are complicated because of the variety of the questions and the different scales used in the answers. This diversity generates certain consistency problems among the different questions and raises issues

[7] Field, 463.

concerning the way the data should be combined for further analysis and interpretation. As a result, before using the factor scores as predictors of usage, it is necessary to conduct a reliability analysis in order to determine the consistency of the developed measurements for the three different factors (liking, nutrition perceptions and awareness). The analysis is supposed to verify that the selected questions measure the same phenomenon, or more broadly speaking, that they are homogeneous.

The most common method to estimate the internal consistency of the questions synthesizing each factor is the Cronbach's alpha test. Cronbach's alpha could provide meaningful results about the homogeneity of selected questions believed to measure phenomena derived from a theoretical frame of reference such as consumer attitudes. The results of the Cronbach's alpha tests for the datasets appear below.

Year	Factor	Alpha	Standardized Item Alpha
1999	Liking	0.8731	0.8679
	Nutrition Perceptions	0.7238	0.7271
	Awareness	0.5705	0.5814
2001	Liking	0.8728	0.8702
	Nutrition Perceptions	0.6741	0.6760
	Awareness	0.6310	0.6385
2003	Liking	0.8465	0.8403
	Nutrition Perceptions	0.6898	0.6908
	Awareness	0.6742	0.6778

To interpret the above results, it is necessary to determine a threshold Cronbach's alpha value. Obviously, higher values for Cronbach's alpha indicate better reliability. Over time, various authors have offered guidelines or rules of thumb regarding minimum levels of acceptable reliability coefficients. In 1967, Nunally stated that reliability coefficients of 0.60 or 0.50 could be considered sufficient for analysis. [8] Nevertheless, in 1978, he modified his preceding statement setting the minimum acceptable level to 0.7. Other authors though agree with Nunally's first assessment. Caplan, Naidu and Tripathi stated that alphas of 0.50 or higher could be judged as adequate for research purposes.[9] Finally, Ellis in 1988 referred to Nunally (1978) and concluded that developed measures could reach acceptable levels of reliability even though several might be lower than 0.7.

The determination of acceptability of a given reliability coefficient is indeed a very important issue. Summarizing the existing background on reliability measurements, it is concluded that standard threshold values of 0.5 or 0.7, even though based on sound reasoning, should always be applied regarding the purposes of the study and the amount of error the researchers are willing to tolerate.[10] In this particular study, the alpha measurements are all well above the threshold of 0.5, and in most of the cases, above or close to 0.7. Consequently, the selected questions and the factor scores derived from them could be considered as acceptable measurements for the developed constructs of liking, nutrition perception and awareness.

4. Relationship Between Attitudes and Usage

The proper method to identify the relationship between the extracted factors and usage is multiple linear regression analysis. This method attempts to fit a predictive model to the existing data and use that model to predict values of the dependent variable, which in this case is usage. Consequently, the goal of the analysis will be to develop a linear equation in which usage is the outcome variable desired to predict and factor scores are the predictors.

The parameters used to assess the validity of the model appear below:

- An F-test will be used to determine whether a significant relationship exists between the dependent variable and the set of the independent variables. This will be referred as a test for the overall significance of the relationship.[11] The test is based on the following hypothesis:

---

[8] Pedhazur, 109,

[9] Pedhazur, 109.

[10] Pedhazur, 110.

[11] Anderson et al., 662.

- H_0: The coefficients of the factor scores all equal 0 $(\beta_1 = \beta_2 = \beta_3 = 0)$.
- H_a: One or more of the parameters is not equal to 0.
- The level of significance used for the test is α=0.05.

- If the F-test shows an overall significance, then t-tests will be used to determine whether each of the individual independent variables (factor scores) is significant. This will be referred as a test for the individual significance. The test is based on the following hypothesis:

For any coefficient of the factor scores $(\beta_1, \beta_2, \beta_3)$,

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

- The goodness of fit for the estimated multiple regression equation will be assessed using the multiple coefficient of determination R^2, which can be interpreted as the percentage of the variability in the dependent variable explained by the estimated regression equation. This coefficient will be adjusted for the number of the independent variables used in order to have a measurement that cannot be affected by adding or keeping non-statistically significant independent variables in the model.

The results of the regression analysis for the three datasets (1999, 2001, and 2003) appear independently below.

Dependent variable: uf

Model Summary			
R	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
0.031	0.001	**-0.003**	2.1503

ANOVA					
	Sum of Squares	**Df**	**Mean Square**	**F**	**Sig.**
Regression	3.176	3	1.059	0.229	**0.876**
Residual	3301.386	714	4.624		
Total	3304.562	717			

Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.641	0.080		20.444	0.000
Liking	0.023	0.081	0.011	0.285	**0.776**
Nutr. Perception	-0.058	0.080	-0.027	-0.721	**0.471**
Awareness	0.024	0.080	0.011	0.303	**0.762**

Table 12. 1999 - Independent Variables: Liking, Nutrition Perception and Awareness for uf.

Dependent variable: unf

Model Summary			
R	R Square	Adjusted R Square	Std. Error
0.074	0.05	**0.000**	2.28

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	14.403	3	4.801	0.923	**0.429**
Residual	2631.472	506	5.201		
Total	2645.875	509			

Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	2.260	0.101		22.327	0.000
Liking	0.00655	0.102	-0.003	-0.64	**0.949**
Nutr. Perception	0.0587	0.101	-0.583	-0.583	**0.560**
Awareness	0.169	0.109	0.069	1.556	**0.120**

Table 13. 1999 - Independent Variables: Liking, Nutrition Perception and Awareness for unf.

Both F-tests above indicate that there is no overall significant relationship between usage as defined by the uf and unf variables and the three identified factors. Also, it is not possible to establish that the coefficients of the regression equations are different than zero conducting the t- test for the individual factors. In interpreting these results, it is possible to conclude that either no relationship between usage and, liking, awareness and nutritional perceptions exists, or that the data used for the analysis are in some manner corrupted.

Dependent variable: uf

Model Summary			
R	**R Square**	**Adjusted R Square**	**Std. Error**
0.214	0.046	**0.042**	2.84986

ANOVA					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Regression	281.153	3	93.718	11.539	**0.000**
Residual	5863.872	722	8.122		
Total	6145.025	725			

Coefficients					
	Unstandardized Coefficients		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	1.81	0.106		17.111	0.000
Liking	-0.474	0.106	-0.163	-4.483	**0.000**
Nutr. Perception	-0.333	0.106	-0.115	-3.151	**0.002**
Awareness	-0.224	0.106	-0.077	-2.117	**0.035**

Table 14. 2001 - Independent Variables: Liking, Nutrition Perception and Awareness of uf.

The results for the uf variable for 2001 are different from those of 1999. The F-test indicates in this case that there is an overall significant relationship between the usage (uf) variable and the liking, awareness, and nutritional perceptions. Also, the testing individually for each of the independent variables provides sufficient evidence to conclude that each is significantly related to the dependent variable (uf). Examining the goodness of fit for the estimated regression equation, it is possible to see that the adjusted R square is extremely small (0.042) explaining only 4.2 % of the total variability of the dependent variable. Therefore, the estimated regression equation is not reliable for use for forecasting.

Dependent variable: unf

Model Summary			
R	R Square	Adjusted R Square	Std. Error
0.116	0.014	**0.008**	3.22

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	71.884	3	23.961	2.315	**0.075**
Residual	5227.256	505	10.351		
Total	5299.139	508			

Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	2.421	0.148		16.391	0.000
Liking	-0.316	0.16	-0.088	-1.977	**0.049**
Nutr. Perception	-0.25	0.139	-0.079	-1.793	**0.074**
Awareness	-0.112	0.156	-0.032	-0.719	**0.472**

Table 15. 2001 - Independent Variables: Liking, Nutrition Perception and Awareness for unf.

Using unf as the dependent variable for 2001, and testing for the overall relationship with the three factors, it was not possible to conclude that it is significant. The contradiction between the results obtained from regressing with uf and unf variables could lead to the conclusion that by excluding the non frequent buyers from the sample, some information is lost or that a type I error[12] has been committed in the F-test for the uf variable.

[12] The error of rejecting the null hypothesis when it is true.

Dependent variable: uf

Model Summary			
R	**R Square**	**Adjusted R Square**	**Std. Error**
0.29	0.084	**0.08**	2.43658

ANOVA					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Regression	374.242	3	124.747	21.012	**0.000**
Residual	4072.72	686	5.937		
Total	4446.961	689			

Coefficients					
	Unstandardized Coefficients		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	2.259	0.093		24.353	0.000
Liking	-0.619	0.093	-0.243	-6.65	**0.000**
Nutr. Perception	-0.265	0.093	-0.104	-2.855	**0.004**
Awareness	-0.304	0.092	-0.12	-3.291	**0.001**

Table 16. 2003 - Independent Variables: Liking, Nutrition Perception and Awareness for uf.

The results from the F-test for 2003 showed that there is an overall significant relationship between the uf variable and the three factors. Also, according to the results of the t-tests, the relationship between uf and each factor individually are found to be significant. Nevertheless, the fit of the regression equation is still very low, explaining only 8% of the variability of usage (unf) (adj. R^2=0.08), and therefore, the model is not reliable for further estimates.

Dependent variable: unf

Model Summary			
R	**R Square**	**Adjusted R Square**	**Std. Error**
0.215	0.046	**0.041**	2.53

ANOVA					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Regression	176.35	3	58.783	9.164	**0.000**
Residual	3656.452	570	6.415		
Total	3832.801	573			

Coefficients					
	Unstandardized Coefficients		**Standardized Coefficients**	**t**	**Sig.**
	B	**Std. Error**	**Beta**		
(Constant)	2.591	0.108		24.042	0.000
Liking	-0.527	0.122	-0.177	-4.312	**0.000**
Nutr. Perception	-0.27	0.107	-0.103	-2.516	**0.012**
Awareness	-0.211	0.111	-0.078	-1.897	**0.058**

Table 17. 2003 - Independent Variables: Liking, Nutrition Perception and Awareness for unf.

The results from the regression analysis of the unf variable for 2003 are more consistent with those of the uf variable. Examining the results of the conducted tests, notice that there is an overall significant relationship between the dependent variable unf and the three factors (F-test is significant p-value =.000). On the other hand, the t-tests indicated that while liking and nutritional perceptions are significantly related with usage (unf), awareness is not at α=0.05 level of significance (p-value = 0.058 > 0.05). Finally, the value of 0.041 for the adj. R^2 is still very low to allow for the safe use of the regression equation for forecasting.

In conclusion, the three datasets (1999, 2001, 2003) did not provide clear evidence about the relationship between usage and consumer liking, nutrition perceptions and awareness. The results for the uf variable seem to be more consistent over time. Indeed, in two out of three datasets examined (2001, 2003), the finding was that there was a significant relationship between the uf and its predictors (liking, awareness and nutritional perceptions). This result, along with the rationale that usage (frequency of purchases) should be at least slightly correlated with some of the three factors, supports the initial hypothesis that the 1999 dataset was corrupted.

Concerning the unf variable results, a significant relationship was identified only in the 2003 dataset. Keeping in mind that the 1999 dataset cannot be considered safe for analysis, and also that the 2001 regression analysis resulted in rejecting the hypothesis, there was an overall significant relationship between usage and independent factors only by 2.5 percent (p-value = 0.075 vs. $\alpha=0.05$). It is possible to conclude that either the variable is biased having excluded the non-frequent buyers (purchases less than once a month), or that the survey process gets more reliable over time.

As far as the conclusion in the 2003 dataset of no significant relationship between the awareness variable and usage (unf variable, p-value = 0.058), it is possible to state that it is an indication that awareness does not significantly impact usage for the frequent purchasers.

Thus far, the three datasets were examined independently despite the fact that the factor analysis identified the same factors over time. Nevertheless, the regression analysis of the factor scores as determinants of the usage of almonds indicated a very low goodness of fit for the developed model. In an attempt to improve the model, it is also possible to consider the idea of pooling the three datasets into one before proceeding with the analysis, assuming that the factors remain constant over time. Appendix D presents the testing procedure followed to validate this assumption before proceeding with pooling the datasets as well as evidence that pooling is not a valid procedure for continuing the analysis.

III. REGRESSION MODELING

This chapter assesses the impact of different expenditure categories on the demand for almonds and creates a model for evaluating the organization's performance. Currently, the ABC categorizes its expenditures in four different buckets relative to the promotion areas in which the organization is involved: public relations, food industry, advertising and nutrition research. The central issue this chapter addresses is developing a methodology that will allow the organization to use its available data effectively to measure performance. Since research in this area is still in the beginning stages, with all the accompanying problems with respect to the quality of the collected data, some consideration will be given to creating a model that will continuously improve with the accumulation of data over time. This model will allow the organization to evaluate the economic impacts of the aforementioned expenditure categories on the demand for almonds.

The analytical procedure implemented for this purpose is primarily regression analysis combined with seasonality analysis and estimation of lag structures among the variables. The procedure will be executed in two stages. The first stage includes the assessment of whether a statistically significant relationship exists between the different expenditure categories and actual demand for almonds. If such a relationship is established, then the second stage proceeds to the development and validation of a model (regression equation) that measures the relative weight of each expenditure category in terms of effectiveness in increasing the demand for California almonds. In this phase, time series analysis will assess the lag structure among the variables and contribute to the identification of the optimal model (highest adj R^2).

A. DESCRIPTION OF THE PROBLEM

The second chapter presented an analysis of the relationship between consumer attitudes and almond usage. That completed the intermediate stage of the conceptual model below, referring to the latent drivers of consumer behavior that affect the almond market. After concluding that certain implicit factors exist that are able to affect almond usage, the next question raised for the ABC is how to allocate its available resources effectively to maximize the influence on those underlying factors, and consequently, to increase the demand for almonds.

The organization directly finances four different categories of promotional programs, identified as expenditures for public relations, advertising, food industry and nutrition research from its budget. Currently, all decisions concerning the allocation of existing funds to the different programs are based on managerial judgment about the effectiveness of each program. This situation can become overwhelming for the management of the organization because of the degree of uncertainty involved with the dynamic changes of the market. Managers, in general, are supposed to have a sense of the market based on their experience that avails them of the ability to make decisions in exceptional cases, such as a sudden trend change that should be addressed by an immediate strategic response. (e.g., launching a new program). Nevertheless, in routine operations such as the annual planning of promotion programs, they need a managerial tool that could provide an objective reference basis concerning the effectiveness of these programs and allow for better decisions in the future.

As already explained in the introduction, the ABC needs a method to classify its expenditures in terms of effectiveness in changing consumer attitudes. Since currently existing data do not make it possible to relate the organization's promotional expenditures to individual attitudes, the only way to evaluate the effectiveness of the promotional programs is to develop a mathematical model that relates the four independent expenditure buckets to the demand for almonds (phase b in Figure 5). The organization does not conduct surveys to identify consumer views about its promotional programs.

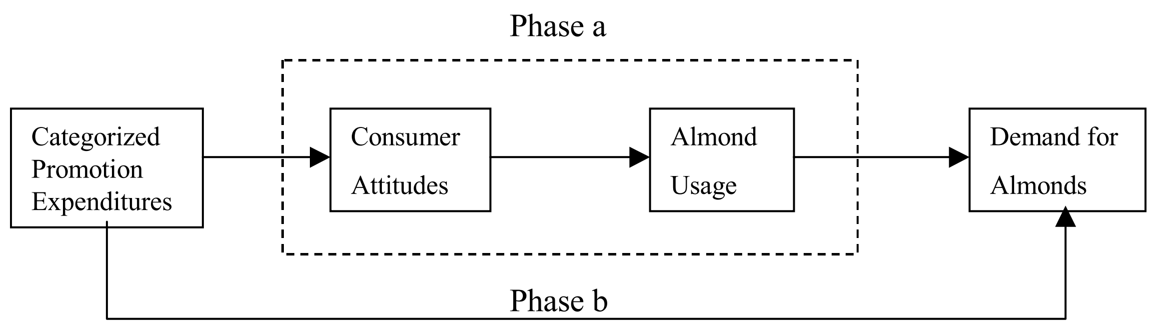


Figure 5. Mathematical Model that Relates the Four Independent Expenditure Buckets to the Demand for Almonds.

This model will present the demand for almonds as a function of the four different expenditure categories (independent variables) in a linear equation. The coefficients of the independent variables in the equation could be used to assess the influence that each expenditure category has on almond demand (dependent variable). Once this model is developed and validated, the ABC management will obtain an effective tool to deal with the existing problem of effectively and efficiently allocating financial resources to its promotional programs in a way that is more consistent with the strategic goals of the organization.

B. DESCRIPTION OF VARIABLES-REQUIRED DATA

The first step in a theoretical approach of creating a model to resolve the aforementioned problem, is properly defining and assessing of the variables that included. From the description of the problem, and considering the needs of the ABC, the authors identified five main variables: public relations, advertising, nutrition research and food industry expenditures will be the independent variables used to predict demand (dependent variable).

1. Public Relations

This variable consists of payments to establish California almonds as a well-known competitive product. The ABC currently has several programs in progress to develop public relations, both in domestic and international markets. These programs include activities in both the United States and several foreign countries, such as announcements of the latest almond nutrition research results, nutrition-focused campaigns, positive TV coverage, public relations with health influencers and participation in international symposiums about nutrition. In summary, the public relations variable includes all those expenditures aiming to build on long-term positive beliefs about almonds.

2. Nutrition Research

This variable refers to expenditures financing research on the nutritional effects of almonds. The ABC has funded various research programs dealing with almond health issues, such as being a premiere source of vitamin E, having the ability to lower cholesterol levels and, in general, containing high value nutrition ingredients.

3. Advertising

This variable includes expenditures for advertisements in mass media. Advertisements mostly include full-page insertions in high circulation food oriented-life style magazines, news ads, and some presence in broadcasting media.

4. Food Industry

Expenditures related to this variable include promotional efforts towards professionals involved in the food manufacturing-preparing market. These efforts include publications in professional magazines and on the web, providing useful information about incorporating almonds in the food service industry, such as almond forms, varieties, nutrition information, and recipes.

5. Demand

Determining a variable able to depict demand accurately is a rather difficult task. Given the fact that demand is defined by a set of two measurements (quantity demanded at a certain price) for a specific time point, in a theoretical model the two measurements could be combined in, one called Sales (=Price*Quantity) in order to be used as the dependent variable. Nonetheless, there are certain timing and record keeping issues associated with the construction of this variable. To capture sales, the ABC should keep detailed records over time for the shipment quantities for all different qualities/ratings of almonds and their respective prices to calculate sales in dollars for the research period.

A second option would be to only use the total shipment quantities of almonds over time as the dependent variable. Shipments definitely present one aspect of demand, indicating the need of the market for almonds or the total quantity that can be absorbed in a certain time period. This is easier in terms of record keeping and employee time requirements, however, it misses the information about pricing of almonds that is valuable for managerial decisions. Despite all that, the information provided in both options for the dependent variable is of major financial importance and extremely useful to the management of ABC. This study will not enter into the argument of selecting one variable over another. The selection decision will reside with the management of the ABC, who will have the necessary information to evaluate the benefits of the model versus the costs to the organization in both cases, making that method the most cost effective choice.

Before describing the methodology of the model development, it is necessary to explain certain adjustments that should be made to the data used in the study. The first concern is about the financial data required for modeling. Knowing that dollar values used to measure expenditures and prices could significantly differ over time due to inflation, especially when the research period extends to many years, it is necessary to adjust all the amounts used in the potential study to a constant year dollar basis using Consumer Price Indexes (CPI) or Producer Price Indexes (PPI). The actual decision of whether or not to implement these adjustments should always be made by the researcher after taking into consideration the quality of the provided data (accuracy in the amounts), the length of the research period, the provided data points for each period and the inflation changes over the period. For example, it would be rather unreasonable to adjust for inflation when examining a short period of years with relatively low inflation, especially when the provided data are on a monthly basis. However, it would be necessary to adjust for longer periods of time with severe inflation fluctuations and extremely accurate data.

The second concern refers to seasonality patterns that could be found in the existing data. Agricultural product shipment quantities and prices usually present cyclical fluctuations over a year's period relevant to the time of the crop season. Examining patterns of shipment quantities over time indicates that there is always a peak after the crop season, while the quantities shipped decrease gradually as the time from that point increases.

When attempting to examine a causal relationship using regression analysis, it is necessary to either isolate the seasonality of the variables (e.g., using seasonal differences or moving averages of the variable's data) that will be entered in the model or in some way integrate it in the model using additional/dummy variables adjusting for the periodicity phenomenon. The choice of the first method over the second depends on the accuracy in determining the seasonality. If the time series analysis results in an extremely accurate equation relating the values of the variable over time, then that equation could be used to isolate the causes of autocorrelation in the variable and extract them before using the variable in the regression analysis. On the other hand, if the seasonality cannot be accurately determined, then the second method is more appropriate. The researcher can simply adjust for the number of different seasons identified in the variable pattern entering n-1 dummy variables in the regression model (where n=the number of seasons identified).

Finally, another issue is the accuracy of determining the time effect the promotion expenditures have on the demand for almonds. The most accurate data the ABC can provide for different categories of promotion expenditures come from ABC accounting books. As a result, the data represent actual payments for the different promotional programs recorded at the time that they were made. The concern with using that data is the possible difficulty to measure the time lag a certain promotional effort has in affecting demand for almonds effectively. The problem occurs because of the time difference between the actual payments and the promotional services.

For example, if a certain service is prepaid, then the available expenditure data refer to the point of the payment, which precedes the actual delivery of services and their effect on demand. As a result, the time lag determined by the model for that specific variable (the one that will provide the best adj R^2 when used as a determinant in the model) will be higher than the actual time lag between the promotional effort enactment and the effect on demand. The opposite happens in case of payments made after the services are provided. Then, the identified lag for the variable in the model is lower than the actual lag. The problem becomes even more complicated if the expenditure data in a certain category for each period consist of payments that are different in nature (a combination of prepayments-payments on delivery and payments after delivery of services). In these cases, determining the real lag structure of the variable is practically impossible. In all other cases, if payments for each expenditure category are fairly homogeneous, market experience can help identify the real lag structure of the variables by simply checking the difference between actual payments and the delivery of services for each category, adding or deducting it from the lag determined in the model, depending on the nature of the payment (prepaid or postpaid).

Another situation, probably more rare, that should also be considered is the existence of a potential lead structure. This could be described as a situation in which the payment for the promotional services is made after delivery but before the effect on the demand for almonds is observed. Figure 6 depicts this sequence of events.

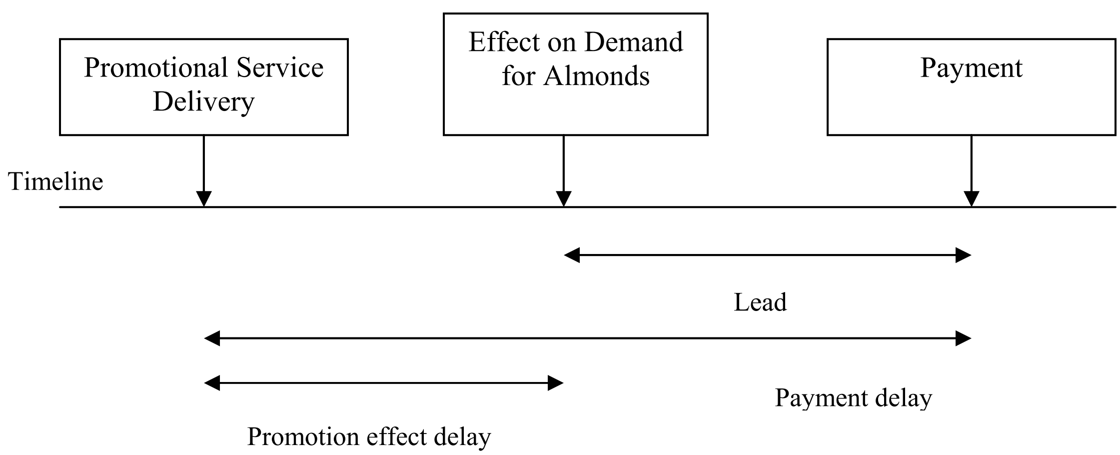


Figure 6. Potential Lead Structure

In this special case, when the time between promotional service delivery and payment is greater than the time it takes the promotional efforts to have an effect and to identify the actual lead, the researcher has to follow a procedure similar to that of the lag structure. Another set of time series variables using the lead function should be created (shifting the existing independent variables backwards by one time unit for as many units as necessary to identify the variable's lead structure) that will be inserted into the regression analysis to determine the best model (max adj. R^2).

Figures 7 and 8 summarize the problem of determining the actual lag structure:

Expenditure paid before the delivery of services

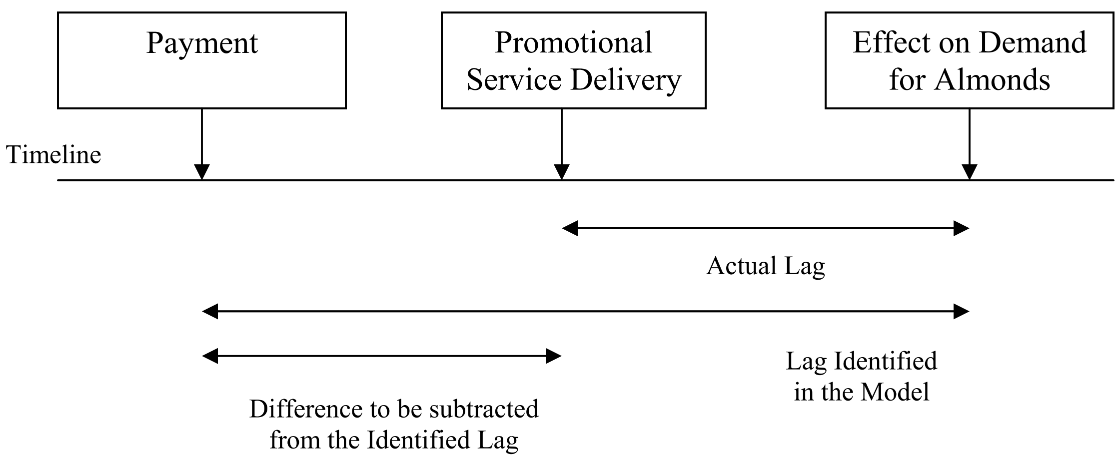


Figure 7. Problem of Determining the Actual Lag Structure

Expenditure paid after delivery of services

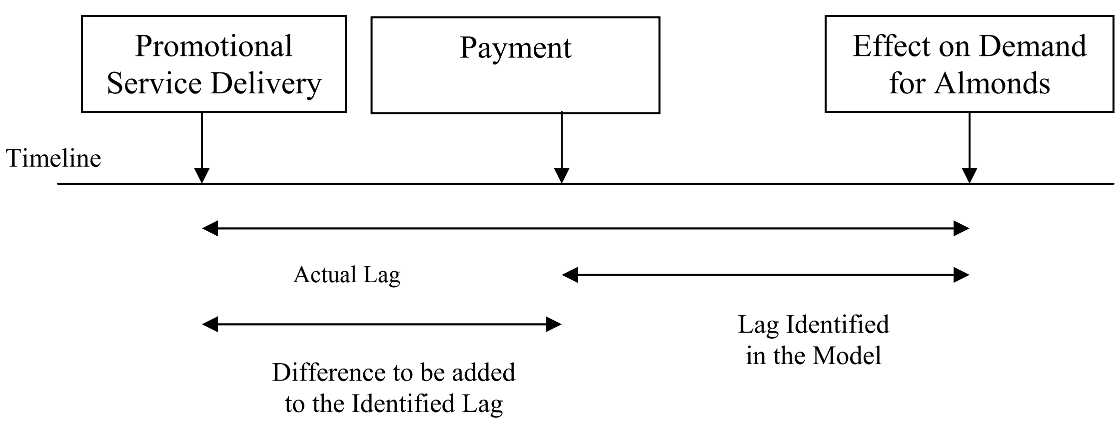


Figure 8. Problem of Determining the Actual Lag Structure

C. METHODOLOGY

As explained in the beginning of the chapter, the method used for creating the model will be multiple regression analysis. A description of the development of a regression model appears as follows

1. Conceptual Model

The conceptual model is related to the actual managerial problem. It analyzes the information ABC decision makers need for better management, combined with assessing the available data on which this analysis is based. In the ABC's case, the requirement was described as a model that relates demand for almonds (sales or shipment quantities) to four promotional expenditure categories (public relations:PR, advertising:ADV, nutrition research:NR and food industry:FI). It is possible to present this requirement mathematically as the identification of the function Equation Chapter 3 Section 3

$$\begin{aligned} &Sales = F\,(PR, ADV, NR, FI) \\ &or \\ &Shipment\ Quantities = F\,(PR, ADV, NR, FI) \end{aligned} \tag{3.1}$$

where sales or shipment quantities are the dependent variables and PR, ADV, NR and FI the independent variables.

2. Examination of the Provided Datasets-Adjustments

Concerns about the datasets that could potentially be used for developing the model were already presented, indicating three main problems: inflation affected figures, seasonality and time lag structures in the datasets. It is necessary to examine the datasets and adjust them properly before using regression analysis in order to manage these problems . Inflation is probably the easiest part to correct, since the adjustment is straightforward. If the researcher decides to adjust for inflation after considering the issues mentioned above, then each expenditure or price amount should be divided by the appropriate price index to adjust for the actual year-month. Evidently, it is necessary for the researcher at this point to determine a base year that will be used as reference for the amounts in constant dollars and for obtaining the necessary CPI numbers. The U.S. Department of Labor/Bureau of Labor Statistics provides detailed data concerning the CPI index on a monthly basis[13] for possible use in transforming values to a common inflation adjusted basis.

Seasonality is a somewhat different problem to resolve. Before proceeding to any adjustments, the researcher has to positively identify seasonal patterns in a variable.

[13] http://www.bls.gov/cpi/home.htm, Accessed May 2004.

There are statistical methods to do this, such as examination of correlation and partial correlations data plots that are beyond the scope of this study. The simplest method, sufficient for this analysis, is to examine a line graph of the data (time series plot) for probable variable changes according to a seasonal regularity. If such patterns are identified, certain adjustments have to be made in either the application of the regression method or to the variables (deseasonalization).

In detail, the researcher has two options; either estimating a seasonality index to isolate the seasonality from the variable before using it in the regression analysis, or adding a number of dummy variables that would control for the periodicity in the regression equation. The selection of a specific method involves scientific judgment. If the researcher believes that seasonality can be effectively eliminated from the variable using the estimated seasonal index, then the deseasonalization method should be implemented. Otherwise, the number of time intervals in each cycle should be determined to create an appropriate number of dummy variables to insert in the regression equation. In general, for k different states (k phases identified in a single cycle) k-1 dummy variables should be created.[14] The method for coding the dummy variables appears below:

Dummy1 : 1 if in phase 1, 0 if else,
Dummy2 : 1 if in phase 2, 0 if else,
Dummy(k-1) :1 if in phase (k-1), 0 if else.

Interpretation of the dummy variables in the regression equation must be considered rather carefully. Negative coefficients (β) for the dummy variables in the regression equation could exist in the regression equation if the high peak is the reference group (the state described by all dummies taking the value of zero). In this case, negative coefficients indicate phases other than peaks, whereas, positive coefficients indicate phases close or at peaks, given that all the dummy β 's are in reference to one group used as the control.

14 Anderson et al., p. 674.

The last issue concerning the data refers to preparing the datasets for identifying time lag structures. The authors previously explained that a delay exists in the effect promotional efforts have on the demand for almonds and expressed concerns about the difficulty in identifying this delay. The procedure to follow in this case is to create a set of time series variables using the lag/lead function (shifting the existing independent variables by one time unit and for as many units as likely to identify the variable's lag/lead structure). These lag or lead variables can be inserted in the regression analysis as independent variables and when the best regression equation is determined (max adj. R^2) it will provide information about the actual existing lag times (payments to effect on demand for almonds).

In any of the above cases, the researcher should pay attention to the sample size before using lead or lag variables as predictors in a regression analysis. Since the model loses as many degrees of freedom as the maximum variable shift, the difference between the sample size and the maximum lead or lag should be large enough to allow for the implementation of the regression method.

3. Implementation of the Regression Method-Assumptions Used-Research Algorithm for Determining the Lag Structure

The main idea behind multiple regression analysis is identifying a linear equation that describes how the dependent variable is related to a set of independent variables. The equation 3.2 presents the general form of a multiple regression model.[15]

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_{12} + \ldots + \beta_n x_n + \varepsilon_i \qquad (3.2)$$

where

y_i is the dependent variable

x_i (i =1,2…n), are the independent variables

ε_i is the error term that accounts for the variability in the dependent variable that cannot be explained by the linear effect of the independent variables.

15 Pindyck Rubinfeld, p. 85.

The application of the multiple regression method presupposes certain assumptions:[16]

- The relationship between the dependent variable and the independent variables is linear and is given by the equation: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_{12} + ... + \beta_n x_n$.
- The independent variables are not stochastic variables. In addition, no exact linear relationship exists between two or more independent variables.
- The error has zero expected value $(E[\varepsilon] = 0)$ for all observations.
- The error term has constant variance $(V[\varepsilon] = c)$ for all observations.
- Errors corresponding to different observations are independent and therefore uncorrelated.
- The error term is normally distributed.

These assumptions are the theoretical basis for implementing of the regression method. As a result, assuring that the data follow these assumptions will validate any developed regression equation.

Regression analysis uses the least squares method to develop the estimated regression equation. The method actually determines the equation that best fits the data minimizing or eliminating the error term.[17] The least squares criterion appears below:

$$\min \sum (y_i - \hat{y}_i)^2 \qquad (3.3)$$

where

y_i = observed value of the dependent variable for the i_{th} observation

$\hat{y}_i$ = estimated value of the dependent variable for the i_{th} observation.

The aforementioned estimated regression equation is valid when a statistically significant relationship exists between the dependent variable and the set of the independent variables. To establish the validity of the regression equation, it is necessary to conduct a series of tests, commonly known as significance tests. Specifically, in

[16] Pindyck Rubinfeld. p. 86.

[17] Anderson et al., p. 647.

multiple regression, the tests used are an F-test and a t-test. The F-test is used to determine whether a significant relationship exists between the dependent variable and the set of all the independent variables (overall significance). If the F-test shows an overall significance, then a t-test is used to determine whether each of the individual independent variables is significant. A separate t-test is conducted for each of the independent variables in the model.[18]

The hypotheses for the tests appear below:

F-test for overall significance:

H_0: The coefficients of the independent variables are all equal to 0 $(\beta_1 = \beta_2 = ...\beta_n = 0)$.

H_a: One or more of the parameters is not equal to 0.

The level of significance used for the test is α=0.05.

t-test for individual significance:

For any parameter β_i

$H_0 : \beta_i = 0$.

$H_a : \beta_i \neq 0$.

In conclusion, the objective of the regression method is to determine the equation that will best fit the data with all the independent variables included in the model significantly related to the dependent variable (F statistic < α = 0.05 and t statistic < α = 0.05 for all coefficients of the independent variables in the equation).

The fit of the model is measured by the multiple coefficient of determination (R^2) that can be interpreted as the proportion of the variability in the dependent variable that can be explained by the estimated multiple regression equation. Since R^2 becomes larger any time a variable is added to the model, even if it is not statistically significant, the

18 Anderson et al., p. 661.

common practice is to use the adjusted multiple coefficient of determination (adj R^2) that compensates for the number of independent variables used in the model, as a goodness of fit measurement.[19]

The actual implementation of the method for the purposes of this analysis is straightforward. The independent variables (PR, ADV, NR, FI) will be inserted into the model in order to assess whether a statistically significant relationship exists between them and the demand for almonds. The main goal of the process is to identify the equation that has the higher possible adj R^2 with all the involved independent variables significantly related to the demand for almonds.

In terms of estimating the time lag that probably exists in the cause and effect relationship (promotional expenditures to market response), identifying the best model becomes a trial-and-evaluate procedure. Assuming that no curvilinear relationships between the variables exists, it might be necessary to run a number of regressions equal to all possible combinations of the existing variables (including shifted variables lead or lag) and assess their results in order to determine the best model.

Given that the lead or lag shifted variables created and tested in the regression are actually determined using subjective methods (manager's intuition about the maximum lag or lead in the variables, researcher's experience etc.), the number of trials varies. For example, in a situation with three independent variables, the researcher believes that the lag structure increases to eight time units (eight additional time series variables created for each one). Then, the following formula calculates the number of possible trials:

$$n = \left(\prod_{i=1}^{k} m\right) - 1 \qquad (3.4)$$

where

n= Number of possible trials,

m=Number of different situations,

k=number of independent variables

19 For the calculation formulas of R^2 and adj R^2, See Anderson et al., pp. 657-658

where the number of different situations = 8+1 (zero lag) + 1 (variable not inserted in the model) = 10 and the number of independent variables = 3. The actual number of possible trials = 10x10x10-1 (non-existent in the model situation for all three variables) = 1,000-1=999.

Since the number of the possible trials increases exponentially as the number of independent variables in the model becomes larger, running the regressions manually is a time consuming procedure. If automation is not possible, then the only route the researcher can follow is to reduce the number of the trials subjectively by combining managerial intuition about the nature of the variables with experience in modeling.

4. Model Validation

Having concluded a regression equation, the next step is to assess its validity. The validation procedure for the constructed regression equation is actually examining of whether the assumptions used in the regression process are correct. With respect to the assumptions presented in the preceding section, the researcher can follow these procedures in order to ensure that no violation has occurred.

First, it is necessary to test the normality of the distribution of residuals by plotting the normal scores of the residuals against the error terms.[20] If the residuals are perfectly normally distributed, then they and the normal scores have the same values. Consequently, combined in the same diagram, they will form a straight line. Therefore, before adopting a final model, this plot should be considered to affirm that no violation concerning the normality of errors assumption has occurred. An alternative approach to test the normality assumption is to construct a histogram of the residuals and examine the symmetry of the distribution. The histogram of the errors should present a pattern similar to a normal probability curve, indicating that the errors are normally distributed.

The next step is to assess the appropriateness of the linear model. By plotting the residuals vs. explanatory variable, it is possible to identify any potential curvilinear relationship in the data. If the plot does not show such a relationship, then the model could be considered appropriate. In the opposite case, it will be necessary to plot the

[20] This is called normal probability plot of residuals. SPSS can generate normal probability plots automatically or when specified by the user.

residuals against all the explanatory variables to determine where the curvilinear relationship originates so that a proper transformation of the related variables can be performed before deriving a valid regression model. The same diagram, (residuals vs. fit plot) also shows probable violations of homoscedasticity (non-constant variance of error terms-heteroscedasticity). The pattern of such a violation is easily identifiable as the error terms expanded as the dependent variable increases.

Finally, to establish the validity of the regression equation, it is necessary to check that the error terms are independent of each other, assuring that no serial correlation exists among successive residual values. A simple graph (line) of the values of error terms over time may easily identify whether this condition exists. If the graph indicates that the independence of error terms assumption has been violated, the model has to be reexamined for potential omissions of important explanatory variables or use of the wrong functional form on the regression equation.

5. ROI Measurement – Financial Interpretation of the Results

The classic definition of Return on Investment (Return On Invested Capital=Net Operating Profit After Taxes/Operating Capital) has to be modified to apply to the ABC case. Given that ABC is a non-profit organization, only interested in measuring the impact that its promotional expenditures have on the demand for almonds, it is then necessary to redefine both profit and operating capital in the above formula.

Assuming that the basic goal of the organization is to maximize the sales of almonds over time, providing additional profit to the almond growers, the profit for the organization should be defined as the positive change in almond sales. Obviously, taxes are not applicable in this concept, and consequently, should be excluded from any calculation. Finally, as operating capital for the organization, the ABC should include the different promotional expenditures, understood as investments for the purposes of the organization. The transformed formula for the ABC ROI is the following:

ROI= (Change in Almond Sales)/ Promotional Expenditures

The only potential usefulness of this formula is as an evaluation tool for the performance of the organization over time. By comparing the ROI from year to year, management can identify overall strengths or weaknesses in its promotional strategy and take corrective actions whenever needed. Nevertheless, the calculation of an overall ROI does not provide managerial information about the effectiveness of the different categories of promotion programs. Therefore, what the organization needs is a specific ROI for each category of expenditures, measured as = Δ (Sales)/Δ(Category Expenditures) indicating the effect on sales for each additional dollar spent in the specific promotion category.

The developed regression models could be used for this purpose. It is possible to derive the ROI for all of them by taking the partial derivatives of the model having identified the relationship between demand for almonds and the different expenditure categories and being able to represent them mathematically in a linear equation.

From equation (3.2), taking the partial derivative of the dependent variable relative to the i_{th} independent variable yields:

$$\partial y / \partial x_i = \beta_i \tag{3.5}$$

(the coefficient of the variable in the developed model) indicating the ROI.

D. APPLICATION EXAMPLE

This section presents the application of the theoretical methodology using the provided datasets from the ABC. Keeping in mind the previously described problems concerning the collection, filtering and limitation of the existing datasets, the authors cannot be confident about the reliability of the potentially developed model. Nevertheless, the application of the methodology itself will allow them to test the flow and robustness of the theoretical analysis, and also, assist any potential researcher in applying the developed method as better quality data becomes available in the future.

1. Available Data-Preparation of the Datasets for Processing

The ABC provided two different files with relevant data for this research. The first file consisted of detailed records for domestic and international monthly shipment quantities from July 1993 to March 2004. The data from this file will be used to form the dependent variable in the analysis. They are the best available source of information about the demand for almonds, given the lack of detailed records about shipments of different quality categories of almonds and their respective prices over time. Despite sales being a better measure of the demand for almonds, shipment quantities can also be used effectively to represent demand for almonds. The numbers provided by the ABC are the actual shipment quantities to domestic and international markets per month. Even though certain seasonality exists due to the nature of the product (an agricultural product with specific crop season and relatively higher number of shipments after that period) these quantities are indicative of the almond market situation, since the shipments refer to the actual demand for almonds. By examining the dataset (presented in Appendix E), notice that an inventory always exists (described as the position of ABC) verifying that there is indeed a difference between produced and shipped quantities of almonds and that the organization has to adjust for the variability in demand using that inventory.

The other file consisted of the monthly promotion expenditures for public relations, advertising, food industry and nutrition research from July 1999 to March 2004. The data were extracted from the accounting records of the organization. As a result, the amounts in the dataset represented the total payments per category for any specific month. Since this data grouping was applied for the first time and due to the difficulty associated with allocating a large number of actual payments to the four expenditure categories, certain omissions were observed, lowering the quality of the actual dataset.

The two datasets are presented in Appendix E. Even though they are presented separately, for the purposes of the analysis they were combined. The time differences between the two datasets resulted in a smaller amount of usable data points for regression (only from July 1999-March 2004).

2. Selection of Variables-Conceptual Model

The purpose of this chapter, as described in the introduction, is to present a methodology for relating the different expenditure categories of the ABC to the demand for almonds. The previous paragraph delineated a clear view of the available datasets for this research. Combining the needs of the ABC management with the existing data constraints, the authors concluded that the goal of developing a regression equation of the following form is:

$$\text{Monthly Shipment Quantities} = \beta_0 + \beta_1 * PR + \beta_2 * ADV + \beta_3 * NR + \beta_4 * FI \quad (3.6)$$

where

monthly shipments quantities (in tons)=Dependent Variable and

PR: monthly public relations expenditures (in dollars)

ADV: monthly advertising expenditures (in dollars)

NR: monthly expenditures for nutrition research (in dollars)

FI: monthly promotion expenditures related with food industry (in dollars) as independent variables.

The regression equation also requires determining the time lag effect that probably exists in the relationship between the expenditure categories and the demand for almonds. The independent variables will be lagged for a certain number of months and inserted in the regression model to be tested for significance. The decision of how many variables will be created will be based on managerial intuition and experience, taking into consideration the number of independent variables required in the model and the existing sample size. All choices concerning lagged variables in the ABC dataset will be explained in detail later. The result, in terms of the model developed, is that the final regression equation will also explain the time lag in the effect each category of expenditures has on the demand for almonds. Its final form resembles the following:

$$\text{Monthly Shipment Quantities } s_t = \beta_0 + \beta_1 * PR_{t-lagPR} + \beta_2 * ADV_{t-lagADV} + \beta_3 * NR_{t-lagNR} \beta_4 * FI_{t-lagFL} \quad (3.7)$$

where lagPR, lagADV, lagNR and lagFI are the potentially identified lags on the variables that create the model with the best fit to the data (max adj R^2 and all independent variables significantly related to the dependent variable). Finally, the coefficients of the independent variables in the equation $(\beta_1, \beta_2, \beta_3, \beta_4)$ would be valuable to the management of the ABC, since they express the weight each expenditure category has on affecting the demand for almonds, or ROI.

3. Assumptions Used-Preparation of the Datasets

With respect to the concerns about the data presented in the theoretical analysis, it is necessary to make the following adjustments.

a. Adjustments for Inflation

The authors chose not to adjust the payments data for inflation. The decision is based on the fact that the available data refer to a relatively short time period with low and stable inflation. Table 18 presents the CPI for the relevant period. Note that the inflation rate was less than 5% for the entire period with small fluctuations that should not significantly affect the amounts used in the analysis.[21]

Year	CPI
1999	2.7%
2000	3.4%
2001	1.6%
2002	2.4%
2003	1.2%

Table 18. CPI

Also keeping in mind that the available data cannot be considered as extremely accurate due to the omission of data points for a period of five months in the middle of the dataset, adjusting for inflation, especially on a monthly basis, is not necessary.

b. Adjustments for Seasonality

Before inserting any variables into the regression model, they were examined or potential seasonality. The time series plots for the independent variables appear in the following figures.

As seen in Figures 9 - 12, the patterns are random, indicating that no serial correlation exists in the independent variables. Consequently, it is not necessary to adjust these variables for seasonality.

[21] The Consumer Price Index (CPI), commonly referred to as the inflation rate, is a measure of the average change in prices paid by consumers for a fixed market basket of goods and services. Inflation has remained low for more than 20 years. Inflation, as measured by the Consumer Price Index has annually remained below 5% since 1991. Source: US Bureau of Labor Statistics, 2004.

Figure 9. Time Series Plots of the Independent Variables for Value Public Relations

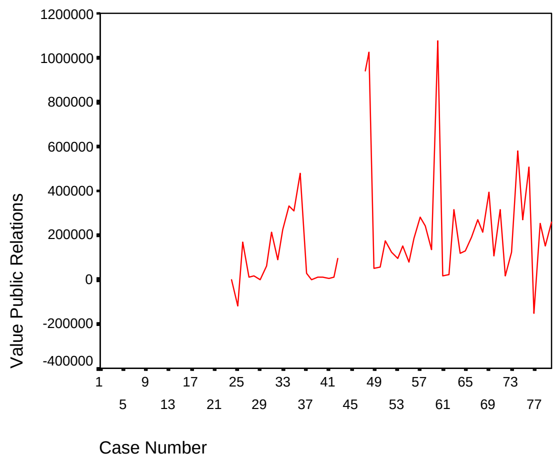


Figure 10. Time Series Plots of the Independent Variables for Value Food Service Industry

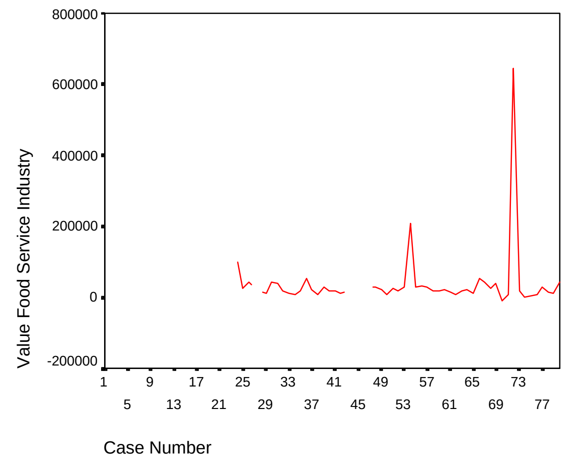


Figure 11. Time Series Plots of the Independent Variables for Value Advertising

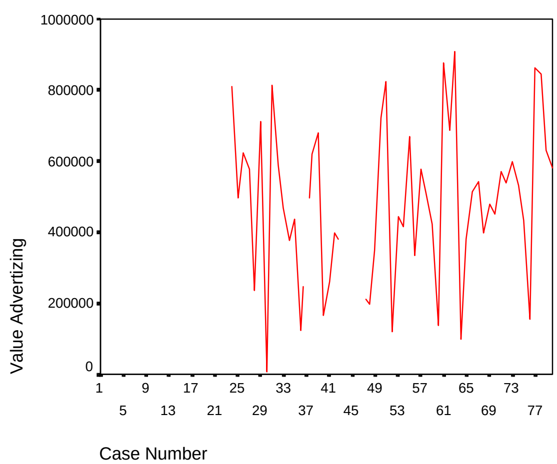
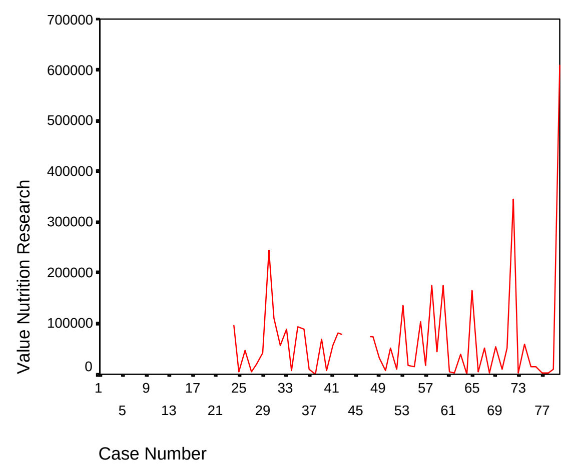


Figure 12. Time Series Plots of the Independent Variables for Value Nutrition Research



The results of the dependent variable's examination are different. As seen in Figure 13, the dependent variable exhibits a repeating pattern over time. Its total cycle equals 12 months, which is the time between two adjacent peaks of the pattern that represent two sequential harvests (agricultural product with fluctuations of constant length). The options at this point are to either eliminate the seasonality from the pattern before using the variable in the regression or account for it in the regression equation. Since there is a great deal of uncertainty involved with accurately decomposing and removing the seasonal pattern, the authors chose to adjust the regression equation for it. The adjustment includes the insertion of certain categorical (dummy) variables in the model representing the identified seasons. These variables will integrate the seasonal regularity in the model, and, as a result, improve its forecasting accuracy.

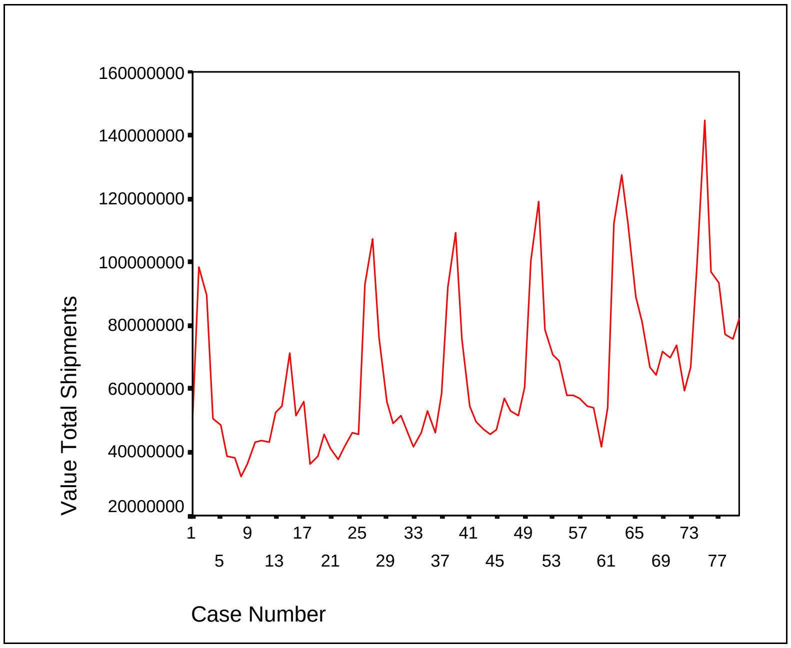


Figure 13. Time Series Plots of the Independent Variables for Value Advertising.

Considering the fact that adding 11 dummy variables (12 months –1) to represent yearly periodicity (examining the same months together) would significantly reduce the available degrees of freedom because of the small sample size, the authors adjusted for seasonality by categorizing data that belong to the same season of the year (Winter, Spring, Summer, Autumn). Three dummy variables using the coding represent the four seasons, presented below:

Dummy 21:1 if shipment month is December, January, February; 0 else

Dummy 22:1 if shipment month is March, April, May; 0 else

Dummy 23:1 if shipment month is June, July, August; 0 else

Autumn, Sept/Oct/Nov, is the reference group. Shipment data in the Autumn months will be represented with 0's in the above three variables. Autumn is exactly after the harvest season that begins in August, and therefore, presents the largest shipment quantities, so the expectation would be to see negative coefficients in the developed regression equation for the three dummies indicating the lower shipments of the other three seasons relative to the peak of Autumn.

c. Assumptions Used in Determining the Underlying Lag Structure

In order to prepare the dataset for determining the time lag in the effect that promotion expenditures have on the shipments of almonds, it is necessary to make two basic decisions:

- whether both lead and lag variables will be created and assessed and,
- what will the maximum number of variables created be, considering the given constraints of the sample size and the available degrees of freedom..

With respect to the first issue previously mentioned, and the question whether a lead structure should be examined, the decision should be based on managerial insights and experience concerning the standard payment procedures of the four promotion expenditure categories and the probability that a certain expenditure category, in which common market practice is to pay after the delivery of services, might have a very fast response time in affecting the demand for almonds. In the ABC case, knowing that a common market practice concerning promotion expenses for advertising, public relations, food industry and nutrition research is to either pay on delivery or even prepay (nutrition research), the authors concluded that it is unlikely for the model to present a lead structure for these variables. Consequently, no lead variables were created.

To determine the number of lag variables created and inserted in the model, the authors decided to test a maximum of one-year lag for advertising, public relations and food industry (12 variables created–sequentially lagging by one month) based on the idea that these expenditure categories respond to the demand for almonds more quickly. For nutrition research, the authors decided to test a maximum of two-years lag (24 variables created-sequentially lagging by one month), given the sample size constraint and understanding that research in general is a more time consuming process with a possible response to the demand for almonds in the long run.

4. Implementation of the Analysis

Having established a conceptual frame for the model, the authors, interested in constructing and concluding the dataset that will be used for its development, proceeded to the analysis part of the research. This part includes the application of the theoretical regression methodology to the dataset as well as the conceptual algorithms used for determining a meaningful/valid model.

Before explaining the procedure followed to reach the model with the best available fit to the data, it is important to present the inherent complexity in the modeling process. Using the formula presented in section 3, the number of different possible models to construct from the existing variables (not including the combinations of dummy variables adjusting for seasonality) in the final dataset is computed as follows:

$$n = (\prod_{i=1}^{k} m) - 1 = 14*14*14*26 - 1 = 71,344 - 1 = 71,343 \quad (3.8)$$

For advertising, public relations and food industry, 14 combinations are possible (No lag+12 possible lags+Not in the model). For Nutrition Research, 26 combinations are possible (No lag+24 possible lags+1 Not in the model).

Despite the fact that certain conceptual assumptions were made for the model, the number of trials was still overwhelming. Even though the number of explanatory variables in the model was limited to a maximum of 4 (excluding dummy variables), each variable was assumed to exist in the model only in one of its possible time lag forms and no curvilinear relationships/transformations of variables or interactions were considered, the trial-and-evaluate procedure for all combinations would be extremely time consuming. Since running and examining the results of all these regressions was almost impossible, even if the procedure was somehow automated, the authors had to determine a method that would allow for tracing the independent variables that were significantly related to the demand for almonds and insert them into the model. Existing statistical variable selection procedures, such as stepwise regression, forward selection, backward elimination could not be used for this purpose, as the number of existing variables to be tested exceeded the sample size (run out of degrees of freedom). Nevertheless, after reducing the number of the potential variables by using a selective (subjective) criterion, it was possible to implement one or more of the aforementioned methods to optimize the model.

Initially, the three created dummy variables were added to the model in order to evaluate their significance. Tables 19 - 21 represent the results.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.719[a]	.516	.497	17210628.48

a. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21

Table 19. Model Summary for Dummy Variables.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.40E+16	3	8.008E+15	27.036	.000[a]
	Residual	2.25E+16	76	2.962E+14		
	Total	4.65E+16	79			

a. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21

b. Dependent Variable: Total Shipments

Table 20. ANOVA Table for Dummy Variables.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.3E+07	3755667		24.841	.000
	DUMMY21	-3.4E+07	5311315	-.621	-6.406	.000
	DUMMY22	-4.2E+07	5449294	-.738	-7.677	.000
	DUMMY23	-4.1E+07	5449294	-.723	-7.516	.000

a. Dependent Variable: Total Shipments

Table 21. Coefficient Table for Dummy Variables.

As seen, the dummy variables added to adjust for seasonality are all significantly related to the dependent variable and alone explain a large proportion of its variability (adj R^2=0.497). This fact raises doubts concerning the contribution of the expenditures to explaining variability in the demand for almonds, something that was expected given that the dependent variable is highly seasonal and not entirely dependant on promotional expenditures. Since the adjustment for seasonality is considered necessary and also explains a large part of the variability in the dependent variable, the authors decided to keep the three dummy variables in the model while tracing the combination of the expenditure variables that provides the best fit to the data.

The tracing method followed was a two-phase process. The first phase consisted of identifying possible predictors/independent variables that, combined with the dummy variables, are significantly related to the demand for almonds. To identify these variables, the authors added them all independently to the basic model consisting of the three dummies and assessed the significance of their relationship with the dependent variable by checking the incremental R^2 of the model versus that of the model including only the dummy variables. The authors used an a=0.15 as a threshold level of significance for the F-test to screen those variables as possible predictors and keep them for further analysis[22]. Appendix F presents the tables indicating the results of this research phase. The following table summarizes the test results concerning the selected variables for the next phase.

22 α =.10 is a default level of significance (used in most statistic software packages) when considering adding (individual p-value being below .10) or removing (individual p-value being above .10) variables from a multiple regression model. The author's selection of 0.15 only allows for more slack in the process- a higher number of variables to be assessed. Nevertheless, the total number of possible combinations is still significantly reduced.

#	Variable	Adj. R^2 w/o Variable	Adj. R^2 with Variable	Change in R^2	F-Sig for change	n
1	Advertising	0.636	0.676	0.044	0.011	52
2	Public Relations Lag 1	0.622	0.651	0.035	0.035	49
3	Public Relations Lag 3	0.661	0.716	0.057	0.004	47
4	Public Relations Lag 6	0.642	0.671	0.035	0.036	45
5	Food Service Lag 3	0.612	0.667	0.058	0.006	48
6	Food Service Lag 5	0.607	0.632	0.031	0.058	46
7	Food Service Lag 8	0.665	0.686	0.027	0.067	43
8	Advertising Lag 8	0.665	0.684	0.025	0.072	44
9	Advertising Lag 12	0.624	0.643	0.027	0.097	40
10	Nutrition Research Lag 2	0.616	0.630	0.021	0.101	50
11	Nutrition Research Lag 3	0.609	0.628	0.025	0.077	49
12	Nutrition Research Lag 15	0.586	0.611	0.034	0.088	39

Table 22. Test Results Concerning the Selected Variables.

After concluding the above set of possible predictors for the final model, the second phase of the process was begun that included inserting all combinations of the selected variables to the regression process and assessing the results for significance and goodness of fit. From the information provided in Table 22, it is possible to calculate the possible combinations for the regression models that include all four variables and will be useful for the purposes of this research.

Advertising: 3 variables selected =3 possible situations.

Public Relations: 3 variables selected =3 possible situations.

Food Service: 3 variables selected=3 possible situations.

Nutrition Research: 3 variables=3 possible situations.

$$\text{Total Trials} = 3 * 3 * 3 * 3 = 81 \tag{3.9}$$

Appendix G presents the regression results for all 81 combinations of the variables. The following criteria were considered to identify the potentially most useful model:

- An overall significance in the relationship between the dependent and the set of independent variables. (F-test significant at α=5%)
- An individual significance for all the independent variables present in the model. (t-test significant at α=5%)
- Maximization of the adj. R^2, indicating the best fit for the model.

Paragraph 6 will present the validation procedure for the selected model.

5. Developed Model–Presentation of the Regression Equation

The trial and evaluate procedure for the 81 combinations resulted in the following model that meets the aforementioned criteria and also does not violate the regression method assumptions.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.932(a)	.869	.818	10103272.39	1.341
a Predictors: (Constant), DUMMY23, Advertising, DUMMY22, LAGS(FOOD_SER,11), LAGS(PUBLIC_R,1), LAGS(NUTRITIO,15), DUMMY21					
b Dependent Variable: Total Shipments					

Table 23. Final Model Summary

Model		Sum of Squares	df	Mean Square	F	Sig.
1	**Regression**	12217468652464160	7	1745352664637737	17.099	.000(a)
	Residual	1837370033124039	18	102076112951335		
	Total	14054838685588200	25			
a Predictors: (Constant), DUMMY23, Advertising, DUMMY22, LAGS(FOOD_SER,11), LAGS(PUBLIC_R,1), LAGS(NUTRITIO,15), DUMMY21						
b Dependent Variable: Total Shipments						

Table 24. Final Model ANOVA

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	73508948.633	6198821.220		11.859	.000
	Advertizing	36.375	10.154	.334	3.582	.002
	LAGS(PUBLIC_R,1)	34.929	9.869	.334	3.539	.002
	LAGS(FOOD_SER,11)	142.163	60.182	.229	2.362	.030
	LAGS(NUTRITIO,15)	125.675	43.584	.308	2.883	.010
	DUMMY21	-34864935.761	5806902.008	-.692	-6.004	.000
	DUMMY22	-41935239.900	6168145.456	-.651	-6.799	.000
	DUMMY23	-51139085.196	6404436.363	-.794	-7.985	.000
a Dependent Variable: Total Shipments						

Table 25. Final Model Coefficients

The developed regression equation formulas appear as equations 3.10, 3.11, 3.12, 3.13):

If t refers to December, January or February for any year:

$$\text{Almond Shipment Quantities}_t = 73508949 - 34864936 + 36.375 * \text{Advertising}_t + 34.929 * \text{Public Relations}_{t-1} + 142.163 * \text{Food Service}_{t-11} + 125 * \text{Nutrition Research}_{t-15} \quad (3.10)$$

If t refers to March, April, or May for any year:

$$\text{Almond Shipment Quantities}_t = 73508949 - 41935240 + 36.375 * \text{Advertising}_t + 34.929 * \text{Public Relations}_{t-1} + 142.163 * \text{Food Service}_{t-11} + 125 * \text{Nutrition Research}_{t-15} \quad (3.11)$$

If t refers to June, July or August for any year:

$$\text{Almond Shipment Quantities}_t = 73508949 - 51139085 + 36.375 * \text{Advertising}_t + 34.929 * \text{Public Relations}_{t-1} + 142.163 * \text{Food Service}_{t-11} + 125 * \text{Nutrition Research}_{t-15} \quad (3.12)$$

If t refers to September, October or November for any year:

$$\text{Almond Shipment Quantities}_t = 73508949 + 36.375 * \text{Advertising}_t + 34.929 * \text{Public Relations}_{t-1} + 142.163 * \text{Food Service}_{t-11} + 125 * \text{Nutrition Research}_{t-15} \quad (3.13)$$

In terms of goodness of fit, the model presented an adj R^2 equal to 0.818, explaining 81.8% of the variability in the dependent variable. Also, from the ANOVA table and the summary of the model, it is possible to observe an overall significance in the relationship between the dependent variable and the set of its predictors (F-test significant – p-value =0.000). Furthermore, all the independent variables in the model are significantly related to the dependent variable at $\alpha=0.05$ level of significance. The p-values for the individual t-tests were Advertising: 0.002, Public Relations: 0.002, Food service industry: 0.030, and Nutrition Research: 0.010.

6. Validation Procedure

The procedure followed to evolve to the model described above consisted of the steps already described in paragraph 4. The first step was testing the normality of the distribution of residuals. From the histogram of the errors (residuals) presented in Figure 14, it is not possible to conclude that the distribution is normal, nevertheless, considering the sample size used to develop the model, the authors cannot positively state there is a violation in the assumption concerning the normality of the residuals.

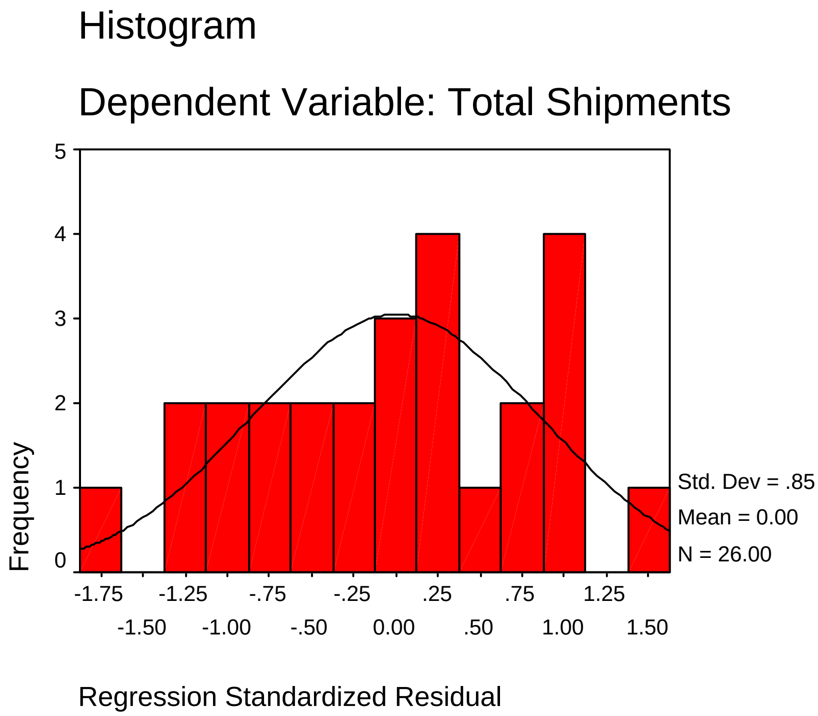


Figure 14. Regression Standardized Residual for Total Shipments.

To support this inference, it is also possible to examine the charts of the normal scores of the residuals against the error terms presented in Figure 15. Based on the chart, notice that there are no significant violations of the normality assumption, as the normal scores of the residuals approach closely the 45 degrees line.

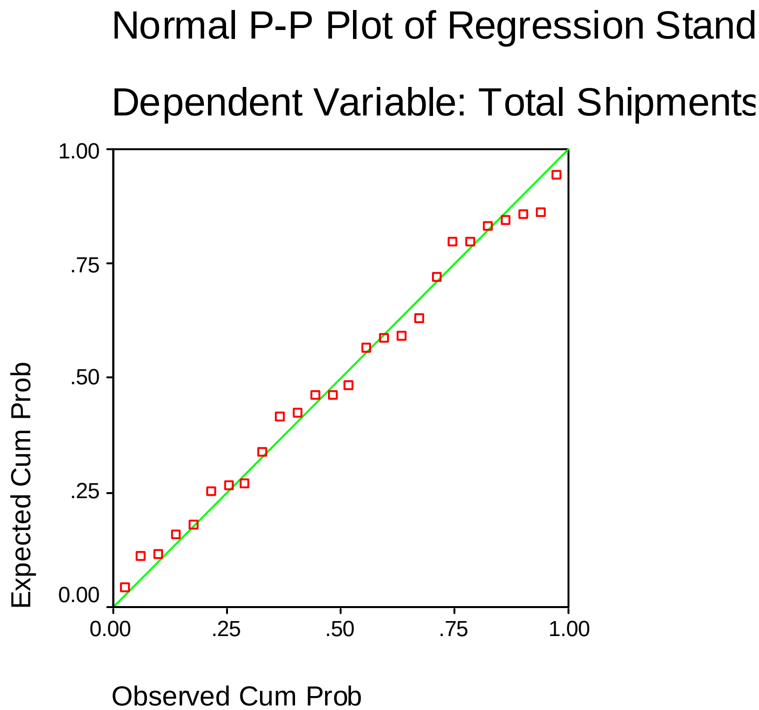


Figure 15. Observed Cumulative Probabilities for Total Shipments.

After verifying that the normality assumption is not violated, the next step is assessing possible homoscedasticity and linearity assumption violations. To establish that the developed model can adequately represent the linear relationship between the variables, it is necessary to examine the residuals vs. fit plot presented in Figure 16.

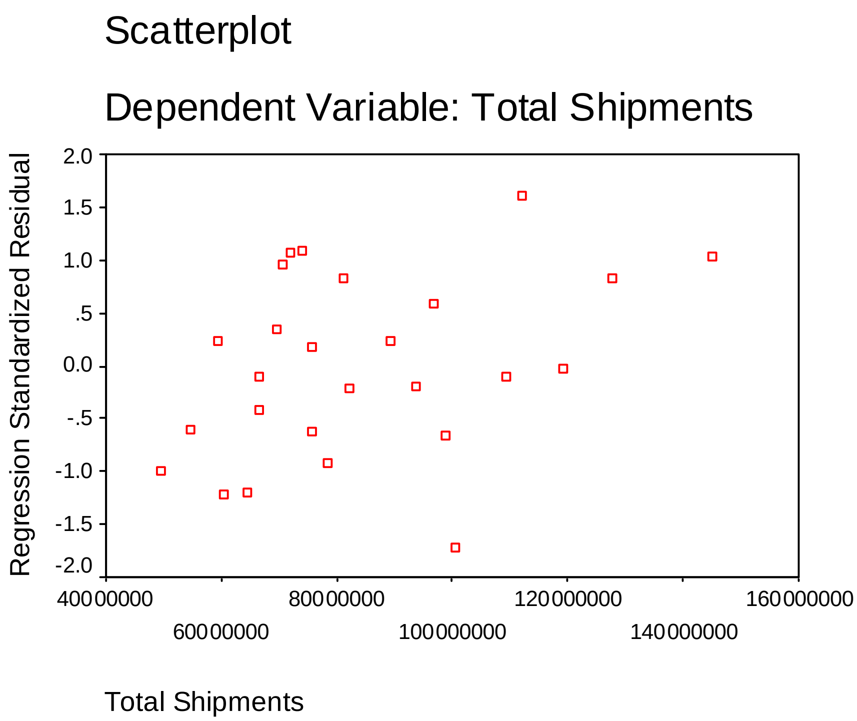


Figure 16. Scatterplot of Standardized Residuals for Total Shipments.

A valid model should not present identifiable patterns in this graph (randomness in errors – some of the residuals above the x=0 line and some below leading to a mean error equal to zero). Observing the above figure, it is possible to conclude that the pattern is close to random or at least that a curvilinear relationship is not present.

Finally, to establish the validity of the developed model, it is necessary have to ensure that the error terms are independent of each other. The graph in Figure 17 presents the unstandardized residuals plotted over time.

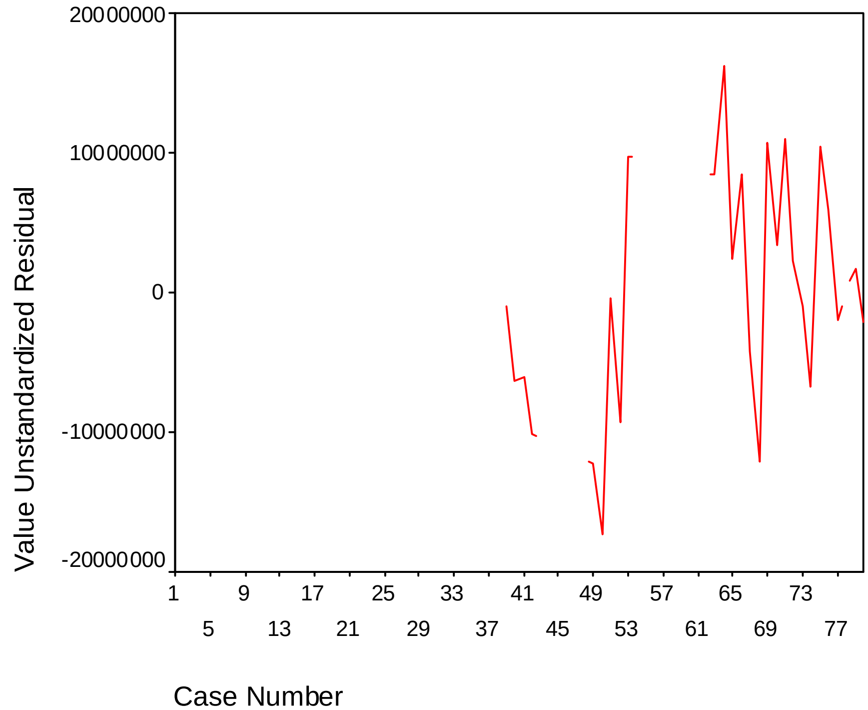


Figure 17. Unstandardized Residuals for Total Shipments.

In general, it was not possible to identify any seasonal patterns in the graph indicating serial correlation in the error terms (violation of the assumption of the error terms being independent). Also, using the Durbin Watson test[23], the authors were not able to identify a correlation problem for the developed model. T Table 26 presents values of the Durbin Watson statistics and the critical values used to test the model

Sample size(n)	Independent Variables	DWstatistic value	D_L	D_U	Conclusion
26	7	1.341	0.64	1.89	Test inconclusive

Table 26. Durbin Watson Statistics and the Critical Values Used to Test the Model.

[23] Durbin – Watson test is a well known and widely used statistical test for first order autocorrelation. It is a summary measure of the amount of serial correlation in the error terms. With uncorrelated errors, the Durbin Watson statistic takes on values near 2. If the errors are perfectly and positively correlated, the D-W statistic will be 0 (for negatively correlated errors the corresponding value is 4). The critical test value depends on both the number of the explanatory variables and the number of the observations in the regression analysis. To determine the limits for significant autocorrelation D-W tables are used that provide the critical values at the 1% significance level . (Summarized from Anderson et al., pp. 729-730).

Given that the Durbin-Watson test is inconclusive, it is not possible to establish that the developed model does not violate the assumption of independent residuals. Therefore, it is necessary to rely on Figure 3.13 and assume the error terms are independent.

7. Interpretation of the Equations – ROI Measurements

The developed model provides relatively clear information concerning the relative magnitudes of the different expenditure categories in affecting almond demand over time, and therefore, could be helpful in determining a fundamental expenditure allocation strategy. From the regression equation, notice that the expenditures for nutrition research could greatly influence the demand for almonds ($\beta = 125.675$) and a probable time lag of fifteen months. For every dollar invested in nutrition research, almond shipments increase by 125.68 thousand pounds 15 months later. This time lag is reasonable, considering that nutrition research projects usually require rather long time periods for completion, and also that the announcement of their results can potentially create market trends. Market reality indicates that marketing campaigns launched in the past, based on results of nutrition research, proved to be extremely effective in positively changing consumer attitudes, developing long term trends towards healthy products. (e.g., low sodium, low cholesterol, low carbohydrates, high protein etc.).

Concerning the food service industry, the model indicates that it has the highest impact on the demand for almonds ($\beta = 142.163$) with a probable time lag of 11 months. For every dollar invested in the food service industry, almond shipments increase by 142.16 thousand pounds 11 months later. This time lag is also reasonable considering that the food industry requires a certain amount of time to increase the consumption of a product (ingredient) by adopting new recipes or changing menus. Furthermore, the benefits from this type of promotion are expected to be considerable, given that the professional market has the ability to absorb large quantities of almonds and significantly influence the existing demand.

The model also presents advertising as having an immediate effect on the demand for almonds (time lag =0), as opposed to public relations expenditures that have a delay of almost one month (time lag =1). The relevant magnitude of both these two expenditure categories is almost the same $(\beta_{adv} = 36.375, \beta_{PR} = 34.929)$, indicating that they have comparable influence on the almond market, significantly lower though than those of the food service industry and nutrition research. This short term effect with relatively lower impact is understandable considering that advertising and public relations focus generally on the individual consumer. In that sense, expect an immediate response from the market is expected, as the influence is directed to the actual consumer basis driving the demand for almonds. However, the actual effect is less significant in terms of magnitude than the major drivers of the market (food service industry, professionals).

The model indicates two different views of the market for managerial decisions: long term vs. short term. The short-term view emphasizesadvertising and public relation expenditures, as opposed to the long term view were food service and nutrition research expenditures are more effective. The incremental improvement of the regression model by adding the variables referring to each view and their importance in explaining the variability in the demand for almonds is presented in Appendix H. The short-term expenditures increase the explanatory power of the model by 19.1% over the seasonality dummy variables. The long-term expenditures explain an additional 10.8% of variability in shipments.

Except for the impact on the demand for almonds, the coefficients that have been previously described provide the additional information in which the ABC is interested: the ROI of the different categories of promotional expenditures. Using the relationship (3.5), and after the derivation, the ROI for each category of expenditures is equal to the β coefficient of each independent variable in the model presented.

In conclusion, the effect of advertising and public relation expenditures is short term and approximately the same in terms of magnitude. On the other hand, the food industry and nutrition research expenditures focus on the long term and could significantly affect the demand for almonds. Considering these findings, the final decision concerning the allocation of the financial resources of the ABC becomes purely strategic as the management of the organization has to evaluate its long term and short term needs as well as the current market conditions and its budget before taking promotional action.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. CONCLUSION AND RECOMMENDATIONS

After careful consideration of the provided datasets, it is possible to draw the following conclusions relative to the initial goals of this study:

The AAU questionnaires that the ABC is currently using have also proved useful in determining consumer attitude factors. From the individual answers to the 1999, 2001 and 2003 surveys, the authors were able to isolate three factors indicating consumer attitudes towards almonds: awareness, liking and nutrition perceptions. Factor analysis allowed for the development of measurements for these factors. The method indicated specific questions from the survey questionnaires possibly to use for calculating factor scores, establishing an objective basis for measuring abstract constructs such as an individual's perceptions about almonds.

With respect to almond usage, the authors concluded that the identified factors have a statistically significant relationship to almond usage as currently measured by the ABC. Nevertheless, the developed regression models from the provided datasets did not prove practical for forecasting purposes. The explanatory power (R^2) was very low, indicating a relatively large amount of unexplained variation in the data. In part, the suspicion is that it is a result of the nature of the dependent variable (almond usage) as opposed to either inadequacy of the identified factors to predict usage for almonds or a potentially inappropriate application of the regression method. ABC currently measures almond usage with a single question relating to frequency of almond purchases in the AAU surveys. Single item scales are difficult to assess for internal consistency or reliability, and purchase frequency is only one aspect of usage. The current method of assessing usage makes it difficult to identify a regression model to predict usage on the basis of consumer attitudes towards almonds. If consumer attitudes towards almonds are assumed to be good predictors of actual almond usage, then focusing on one of its dimensions only (frequency of purchases) could result in a low quality model due to missing information in the dependent variable. Therefore, it appears that better results could be obtained from a regression model if measures of almond purchases were also used to describe usage.

Regarding the quality of the datasets provided, the only concern was about the 1999 survey. Although no problems were found in all three datasets with respect to the sampling procedure, the individual answers of the 1999 dataset did not seem to be consistent with those of the other two surveys. Even though factor analysis indicated that the same questions should be used to identify the factors in all three years, the relationship between almond usage and the three identified factors only proved not to be statistically significant for 1999. This could be due to many different reasons, such as the immaturity of the survey since the AAU surveys was first launched in 1999 and observed changes in both the questions and the way they were stated after 1999,, some type of transition-changes period relating usage and consumer attitudes or even corruption of the dataset. The authors recommend avoiding the 1999 dataset for further research as opposed to the other two surveys, conducted in 2001 and 2003.

The recommendations to the ABC, concerning improving the method for measuring consumer attitudes and establishing a model to forecast actual almond usage using factors such as consumer attitudes are as follows:

- The AAU questionnaires should include non-extremely scientific (common) and more clearly stated questions about beliefs and attitudes relative to almonds. Questions that require specific knowledge about almond issues should be avoided. The scale used for answers should be ratio or interval to be useful for factor analysis. Negatively stated questions using the same interval scale for responses, e.g., 1-5 where one means strongly agree and 5 strongly disagree, should be avoided or restated as they increase ambiguity for the person taking the survey, resulting in lower quality response data. In the research, all these problems were identified during the factor analysis of the data, leading to the exclusion of questions that could potentially provide additional measuring capability to the factor scores. Questions that seemed to fit conceptually to certain factors had to be excluded. They did not correlate well to the others in the same factor because of their higher variability-inconsistency in responses due to extreme specialization.
- The current measure of almond usage should change to something more comprehensive, with multiple items. The question in the surveys referring to purchase frequency for almonds does not cover the issue of actual almond consumption that the ABC is interested in predicting by the estimated individual scores on the attitude factors of Awareness, Liking and Nutrition Perceptions.

- To integrate the two research phases presented in this study, ABC should also consider adding items/questions to the AAU surveys relating to promotion expenditure categories and their ability to influence demand. The responses to these questions will help determine the relationship between the identified factors of awareness, liking and nutrition perceptions and the existing promotion expenditure categories setting the basis to develop a complete almond market assessment model.
- Finally, the organization should consider more frequent AAU surveys. Even though the difficulties associated with regularly launching surveys (time and funding constraints) are well understood, it is important for the ABC to measure consumer attitudes at least once a year to have the data determine changes over time and be able to react to these changes (changes in the promotion strategy). For this purpose, it is possible to develop shorter questionnaires including only the questions identified by the factor analysis, and thus minimizing the implementation-reporting time for the surveys and overall cost of the process.

In the second phase of the research study, the authors attempted to develop a methodology for examining the relationship between promotion expenditures and almond demand. Despite the existing problems concerning the accuracy/quality of the provided datasets, it was possible to draw the following conclusions.

The demand for almonds, as expressed by the variable monthly shipment quantities, reflects high seasonality. This fact raises issues about the appropriateness of the regression analysis to explain or determine a causal relationship between different categories of promotion expenditures and demand for almonds with the available data. The authors were able to explain this 49.7% of the variability in the dependent variable by simply adjusting for seasonality and an additional 32.1% when including expenditure predictors (reaching a total adj. R^2 of 81.8% in the best model).

On the other hand, the results are not discouraging. The authors were able to establish that expenditures are significantly related to the demand for almonds, as well as verify the initial assumption concerning the existence of a time lag structure in the way the promotion expenses influence the consumption of almonds. Despite the problems with the size and the quality of the sample, one final regression model able that provides information about the actual relationship between the different promotion categories and the shipment of almonds was identified. A summary of the findings from the regression analysis are as follows:

- Advertising and public relations expenditures can be considered the most immediate in terms of affecting the demand for almonds, as the time lag between the payment of these expenditures and the effect in demand is close to zero (0 -1 months lag identified in the model). In addition, their relative magnitude of effectiveness is in the range of 35 to 36 (β's or ROI coefficients identified in the model), meaning that for every additional dollar spent in advertising and public relations, the immediate effect is an increase of 35 to 36 thousand pounds in almonds shipments.
- Nutrition research could have a very significant long-term impact on the demand for almonds. The time lag for these expenditures is relatively longer than other categories, (the model identified a lag of 15 months). Nevertheless, the magnitude of the effect nutrition research has on the demand for almonds is one of the highest among all categories. The developed models indicated a β coefficient of 125, meaning that for every additional dollar spent on nutrition research, an increase of almost 125,000 pounds is expected in the almond shipments after 15 months.
- Food service expenditures have a relatively shorter time lag in affecting the demand for almonds compared with nutrition research. The developed model indicated that promotion expenditures in the food service industry are expected to affect the market after a period of 11 months. Its effect on shipments could be considerable. The β coefficient of the estimated regression equation was approximately 142 indicating the largest magnitude relative to the other expenditure categories.

Accepting the statistical error risk in the validity of the developed model, two options are implied concerning the ABC strategy. The first focuses on the long run, emphasizing on nutrition research and food services industry, providing potentially higher yields in terms of impact on demand for almonds after a period of 11-15 months. The other option is considered short term. It recommends taking promotional action aimed at advertising and public relations, for lower yields in terms of raising demand, nevertheless, in shorter time (0-1 months delay).

Keeping in mind those two options, the authors believe that the choices concerning the allocation of the financial resources should be made using simple rules. The development of the regression model allows ABC to have an objective function for optimizing its portfolio of investments/promotion expenditures. From that point forward, achieving an effective and efficient promotional strategy for the organization simply becomes an issue of setting the proper constraints. In terms of funding, this task is simplified because constraints are already set by the organization's budget. However, the market is a dynamic environment that changes continuously. Determining the effect that promotion could have on consumer behavior remains a challenge that will have to be continuously addressed by the ABC management.

Finally, to improve the methodology for determining the effect that different categories of promotion expenditures have on the demand for almonds, the ABC should consider the following:

- The organization should keep detailed monthly data about almond shipment quantities, including all different categories of almonds and their respective prices over time. Thus, quantities shipped can be matched with respective prices to determine actual sales, a dependent variable that could better represent demand for almonds in a regression model.
- The management should implement a standardized procedure for extracting expenditure data from the organization's accounting records and categorizing them in "buckets" for which there is a managerial interest. This procedure should be continued over time, providing the necessary datasets for statistical research. In addition, the organization should collect data concerning the nature of payments and current market practices to identify of the time lag structure that the promotion efforts have in affecting the almond market.

Adopting the above recommendations, the organization will be able to obtain higher quality data that would improve the accuracy and reliability of the developed regression model. After all, measuring the return on investment for the different promotional expenditure categories of the ABC proved to be a challenge due to the existing uncertainty in the given datasets. Consequently, improving the quality of the datasets allows for a better study of almond market behavior and provides the necessary inputs to adjust the developed model. Thus, the ABC will maintain its ability to understand the factors that drive the demand for almonds over time, allowing for better decisions concerning almond promotion in the domestic and international markets.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. QUESTIONS USED FROM THE AAU QUESTIONNAIRE

Question 1: -*When you think of nuts, which nut comes to mind FIRST?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was Almonds
Value=0 if else.

Question 2: - *Thinking about nuts in general, I would like you to rate how well you like them. Using a scale from 1 to 10, when 1 means you don't like nuts at all and 10 mean you like nuts extremely well, which number best represents your opinion?*

The question refers to nuts in general. The provided data (answers) were recoded using the exact opposite scale: 1 meaning that one does like nuts extremely well and 10 meaning that one does not like nuts at all. The purpose of the above mentioned transformation is to have a unidirectional scale of answers with lowest numbers indicating positive attitudes towards almonds/nuts versus larger numbers indicating negative attitudes.

Question 3: - *Offhand, within the past year, can you recall seeing or hearing any advertising for nuts of any type?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Yes"
Value=0 if else.

Question 4: - *Please tell me which nut is best described by the statement –"the special ingredient that makes other foods better"*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Almonds"
Value=0 if else.

Question 5: - *Please tell me which nut is best described by the statement – "the special nut that eats well as a snack and helps my heart"*
The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Almonds"
Value=0 if else.

Question 6: - *Which type of nut, if any, is your favorite?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Almonds"
Value=0 if else.

Question 7: - *Have you seen or heard any advertising within the past year for … almonds?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Yes" for "Almonds"
Value =0 if else.

Question 8 - *Excluding advertising, within the past year, can you recall reading any articles or hearing a health story about nuts in any public media, such as newspapers, magazines, radio, TV or online?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Yes"
Value=0 if else

Question 9 - *Which types of nuts were mentioned in the articles or health stories?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Almonds"
Value=0 if else

Question 10 - *When you think about nuts that are a part of candy, what one type of nut FIRST comes to mind?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Almonds"
Value=0 if else

Question 11 - *When you think about nuts that are a part of ice cream, what one type of nut FIRST comes to mind?*
The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:
Value=1 if the answer to the question was "Almonds"
Value=0 if else

Question 12. *When you think about nuts that are a part of a cereal, what one type of nut FIRST comes to mind?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:

Value=1 if the answer to the question was "Almonds"

Value=0 if else

Question 13. *Now, think for a moment about nuts as a snack by themselves, as a whole nut. When you think about eating nuts as a snack, what one type of nut FIRST comes to mind?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:

Value=1 if the answer to the question was "Almonds"

Value=0 if else

Question 14. *What nuts do you choose to eat, either by themselves or in other foods, on an ongoing basis?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:

Value=1 if the answer to the question was "Almonds"

Value=0 if else

Question 15. *Within the past year do you recall using any recipes from magazines or newspapers that called for nuts?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:

Value=1 if the answer to the question was "Yes"

Value=0 if else

Question 16. *And what type of nut or nuts did you use as that ingredient for a recipe or to add to a food dish you prepared?*

The question refers to nuts in general. For analytical purposes the provided data (answers) were transformed as follows:

Value=1 if the answer to the question was "Almonds"

Value=0 if else

Question 17. *Thinking about almonds in general, and using the 1 to 10 scale where 1 means you don't like almonds at all and 10 means you like almonds extremely well, which number best describes your opinion?*

The question refers directly to almonds. For conformity purposes (lower values indicating positive attitudes for almonds and higher values negative ones), the provided data were recoded on the exact opposite scale. (1 indicating the highest positive attitude and 10 the highest negative attitude)

Question 18. *Thinking specifically about the taste of almonds, on a 1 to 10 scale, where 1 means you don't like the taste at all and 10 means you like the taste extremely well, which number best describes your opinion?*

The question refers directly to almonds. For conformity purposes (lower values indicating positive attitudes for almonds and higher values negative ones), the provided data were recoded on the exact opposite scale. (1 indicating the highest positive attitude and 10 the highest negative attitude)

Question 19. *Thinking specifically about the crunchiness of almonds, on a 1 to 10 scale, where 1 means you don't like the crunchiness at all and 10 means you like the crunchiness extremely well, which number best describes your opinion?*

The question refers directly to almonds. For conformity purposes (lower values indicating positive attitudes for almonds and higher values negative ones), the provided data were recoded on the exact opposite scale. (1 indicating the highest positive attitude and 10 the highest negative attitude)

Question 20. *Thinking specifically about the healthfulness of almonds, on a 1 to 10 scale, where 1 means you don't think almonds are healthy at all and 10 means you think almonds are very healthy, which number best describes your opinion?*

The question refers directly to almonds. For conformity purposes (lower values indicating positive attitudes for almonds and higher values negative ones), the provided data were recoded on the exact opposite scale. (1 indicating the highest positive attitude and 10 the highest negative attitude)

Question 21. *Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds is more interesting.*

The raw answers were provided in the following form:

Value 1 =Strongly Agree
Value 2 =Somewhat Agree
Value 3=Somewhat Disagree
Value 4=Strongly Disagree
Value 5=D/K N/A

In order to have a meaningful scale-more consistent to that of other questions (1-10 scale with 1 indicating the highest positive attitude), the rating 1 to 4 was maintained with the exception of the answer D/K-N/A that was recoded taking a value of 2.5 (the middle of the scale) instead of the value 5, as it indicates an indifferent attitude to the question's statement.

Question 22. *Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds is special to serve to family or friends*

The raw answers were provided in the same form as question 21. Consequently, the following recoding was applied:

Value 1 =Strongly Agree
Value 2 =Somewhat Agree
Value 2.5= D/K – N/A
Value 3=Somewhat Disagree
Value 4=Strongly Disagree

Question 23. *Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds is worth more money*

The raw answers were of the same format as in question 22. The same recoding as in question 22 was also applied.

Question 24. *Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds is worth more money is contemporary.*

The raw answers were of the same format as in question 22. The same recoding as in question 22 was also applied.

Question 25. *Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds is better nutritionally.*

The raw answers were of the same format as in question 22. The same recoding as in question 22 was also applied.

Question 26. *Would you say you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement; "almonds are comparable to fruits in their ability to positively affect your health"?*

The raw answers were of the same format as in question 22. The same recoding as in question 22 was also applied.

Question 27. *Thinking only about cholesterol and almonds, do you think almonds are high, moderate or low in cholesterol, or do almonds contain no cholesterol at all? If you aren't sure, just say so.*

The raw answers were provided in the following form:

Value 1 =High
Value 2 =Moderate
Value 3=Low
Value 4=Not at all
Value 5=D/K N/A

In order to have a more consistent scale to that of other questions and knowing that the belief that almonds are high in cholesterol indicates a negative attitude to almonds, the following transformations were applied:

The rating 1 to 4 was reversed as presented below and the answer D/K-N/A was recoded taking a value of 2.5 (the middle of the scale) instead of the value 5, as it indicates an indifferent attitude to the question's statement.

Value 1 = Not at all
Value 2 =Low
Value 2.5=D/K N/A
Value 3= Moderate
Value 4= High

Question 28. *Using ratings of excellent, good, fair, or poor, how would you rate almonds for being Nutritious*

The raw answers were provided in the following form:

Value 1 =Excellent
Value 2 =Good
Value 3=Fair
Value 4=Poor
Value 5=D/K N/A

In order to have a meaningful scale and more consistent to that of other questions (1-10 scale with 1 indicating the highest positive attitude), the rating 1 to 4 was maintained with the exception of the answer D/K-N/A that was recoded taking a value of 2.5 (the middle of the scale) instead of the value 5, as it indicates an indifferent attitude to the question's statement.

Question 29. *Using ratings of excellent, good, fair, or poor, how would you rate almonds for being Low in saturated fat?*

The raw answers were of the same format as in question 28. The same recoding as in question 28 was also applied.

Question 30. *Using ratings of excellent, good, fair, or poor, how would you rate almonds for being High in vitamin E?*

The raw answers were of the same format as in question 28. The same recoding as in question 28 was also applied.

Question 31. *Using ratings of excellent, good, fair, or poor, how would you rate almonds for being high in protein?*

The raw answers were of the same format as in question 28. The same recoding as in question 28 was also applied.

Question 32. *Using ratings of excellent, good, fair, or poor, how would you rate almonds for being heart healthy?*

The raw answers were of the same format as in question 28. The same recoding as in question 28 was also applied.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. EXPLORATORY FACTOR ANALYSIS FOR THE YEARS 2001 AND 2003 DATA

Examining the output of the analysis, it is possible to conclude that the developed model for 1999 is also valid for 2001 and 2003. The selected questions/variables correlate fairly well with each other as indicated in the R-significance level matrices, and consequently, it was not necessary to exclude any variables from the models. As a result, the three models (1999, 2001, 2003) included the same variables/questions. The determinants of the R-matrices were 0.0172 and 0.0197 for 2001 and 2003 respectively, showing that there were no multicollinearity issues in the data.

Regarding sampling adequacy, that authors concluded that both years' data were appropriate for factor analysis. The KMO values of 0.818 and 0.808, the greater than 0.5 diagonal values and the low (close to 0) off diagonal elements of the anti-image correlation matrices, point to the same direction as well.

The Bartlett's test of Sphericity is also highly significant (Chi square = 2963.812, df=55, Significance=0 for 2001 and Chi square = 2726, df = 55, Significance =0 for 2003) confirming that relationships exist among the variables that will be included in the analysis.

Using the same factor extraction process, three factors have been determined for 2001 and 2003, explaining 47.491 % and 48.607 % of the total variance respectively. The scree plots, even though they cannot be easily interpreted, imply that the number of factors that should be retained is between two and four. Consequently, the decision to retain three factors seems to be acceptable for the purpose of this analysis.

Year 2001

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.919	35.625	35.625	3.536	32.144	32.144	2.602	23.654	23.654
2	1.597	14.522	50.147	1.050	9.544	41.689	1.358	12.346	36.000
3	1.201	10.923	61.070	.638	5.802	47.491	1.264	11.491	47.491
4	.897	8.153	69.223						
5	.679	6.173	75.396						
6	.642	5.833	81.229						
7	.608	5.531	86.761						
8	.509	4.627	91.388						
9	.487	4.425	95.813						
10	.349	3.177	98.990						
11	.111	1.010	100.000						

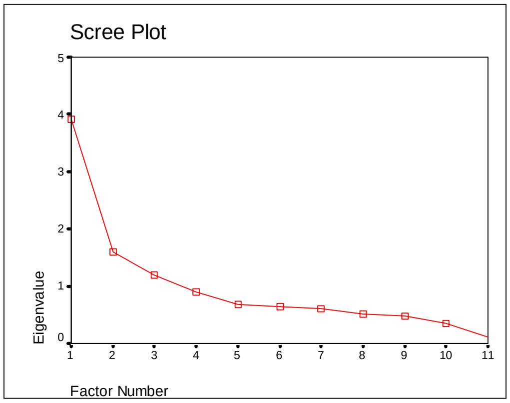
Extraction Method: Principal Axis Factoring.

Year 2003

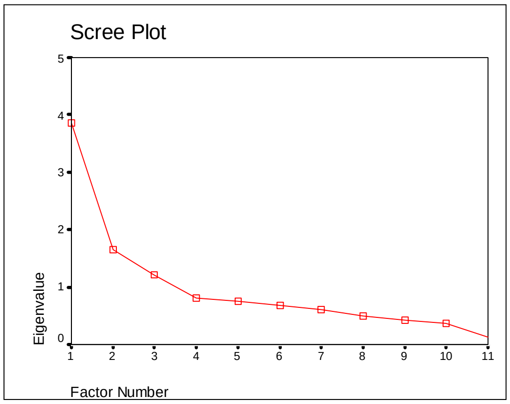
Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.864	35.123	35.123	3.481	31.645	31.645	2.429	22.085	22.085
2	1.659	15.078	50.201	1.161	10.550	42.195	1.557	14.153	36.239
3	1.210	10.996	61.197	.705	6.413	48.607	1.361	12.369	48.607
4	.806	7.326	68.523						
5	.749	6.812	75.335						
6	.682	6.201	81.536						
7	.611	5.556	87.091						
8	.503	4.572	91.664						
9	.418	3.796	95.460						
10	.369	3.354	98.814						
11	.130	1.186	100.000						

Extraction Method: Principal Axis Factoring.



2001



2003

The results of the factor analysis for all data sets are presented below. As seen in the table, the same variables/questions were grouped under the same factors. Also, there are only slight differences in factor loadings after rotation among the three datasets, proving that the developed models are consistent over time and that the selected question groups could be used for measurement of the identified three factors.

Rotated Factor Matrices									
	1999			**2001**			**2003**		
	Factor								
Question	**Liking**	**Aware**	**Nutr. Perc.**	**Liking**	**Aware**	**Nutr. Perc.**	**Liking**	**Aware**	**Nutr. Perc.**
2	0.462			0.525			0.429		
17	0.921			0.910			0.927		
18	0.914			0.902			0.860		
19	0.693			0.700			0.663		
21		0.698			0.628			0.664	
22		0.662			0.586			0.620	
23		0.549			0.559			0.557	
28			0.729			0.744			0.851
29			**0.352***			0.427			0.514
30			0.418			0.498			0.498
31			0.507			0.502			0.460

This analysis has revealed that we are able to extract certain questions/variables from the survey used by the ABC that could provide measurements for individuals' attitudes towards almonds. In reality, the developed model can effectively measure the constructs of "liking", "awareness" and "nutritional perceptions" using the scores of the extracted factors.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. SCALES EMERGING FROM THE EXPLORATORY FACTOR ANALYSIS

A. LIKING QUESTIONS

2. Thinking about nuts in general, I would like you to rate how well you like them. Using a scale from 1 to 10, when 1 means you don't like nuts at all and 10 mean you like nuts extremely well, which number best represents your opinion?

17. Thinking about almonds in general, and using the 1 to 10 scale where 1 means you don't like almonds at all and 10 means you like almonds extremely well, which number best describes your opinion?

18. Thinking specifically about the taste of almonds, on a 1 to 10 scale, where 1 means you don't like the taste at all and 10 means you like the taste extremely well, which number best describes your opinion?

19. Thinking specifically about the crunchiness of almonds, on a 1 to 10 scale, where 1 means you don't like the crunchiness at all and 10 means you like the crunchiness extremely well, which number best describes your opinion?

B. AWARENESS QUESTIONS

21. Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds *is more interesting.*

22. Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds *is special to serve to family or friends*

23. Next I will read several statements that could be used to describe products that contain almonds. After I read each one, please tell me if you strongly agree, somewhat agree, somewhat disagree or strongly disagree with the statement. First, a product that contains almonds *is worth more money*

C. NUTRITIONAL PERCEPTIONS' QUESTIONS

28. Using ratings of excellent, good, fair, or poor, how would you rate almonds for being Nutritious.

29. Using ratings of excellent, good, fair, or poor, how would you rate almonds for being Low in saturated fat?

30. Using ratings of excellent, good, fair, or poor, how would you rate almonds for being High in vitamin E?

31. Using ratings of excellent, good, fair, or poor, how would you rate almonds for being high in protein?

APPENDIX D. POOLING CONSIDERATIONS

A valid procedure to test the project's assumption for unchanged factors over time should be to compare the means of the factor scores for liking, nutritional perceptions and awareness, among the three datasets. Since the surveys were conducted using different samples, an independent means' t-test based on the following hypothesis would be appropriate to analyze the situation:

$$H0 : \mu 1 - \mu 2 = 0$$
$$Ha : \mu 1 = \mu 2 \neq 0$$

μ_1, μ_2 are the means of the factor scores for each variable for the first and the second dataset respectively. This test should be performed nine times (three times for each factor 1999 vs. 2001, 1999 vs. 2003, 2001 vs. 2003 multiplied by the three existing factors).Due to the available calculation methods for the factor scores (Anderson Rubin method, regression method) provide only standardized results (mean = 0). It is necessary to calculate non-standardized factor scores before proceeding with the hypothesis test described above.

The following equation describes the calculation method for the non-standardized factor scores:

$$FS = \sum (Q*F)$$

FS is the Non–standardized factor score, Q is the individual's response for each question and F is the factor score coefficient for each selected question and respective factor.

The result of this calculation is a non-standardized factor score for each individual in any sample that allows for the comparison between the factor score means of the three yearly datasets. After calculating the factor scores for all three datasets, the comparison of their means using an independent samples t-test as described above occurred next. The results of the nine conducted tests are the following:

- For factor 1 (Liking)

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
1999	733	**3.560299**	2.701613	0.0998
2001	735	**3.532833**	2.56536	0.0946

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
0.2	1466	**0.842**

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
1999	733	**3.560299**	2.701613	0.0998
2003	700	**3.375448**	2.697645	0.101961

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
1.296	1431	**0.195**

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
2001	735	**3.532833**	2.56536	0.0946
2003	700	**3.375448**	2.697645	0.101961

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
1.133	1433	**0.257**
1.131	1419.094	**0.258**

The calculated p-values of 0.842, 0.195 and 0.257 for the 1999 vs. 2001, 1999 vs. 2003 and 2001 vs. 2003 datasets respectively, well above the $\alpha = 0.05$ level of significance, indicated that we could not reject the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ and conclude that the means of the factor scores between the three datasets are different. Consequently, proceeding with pooling the three datasets into one could only be conducted by taking the risk of a type II error.[24]

- For factor 2 (Nutritional Perceptions)

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
1999	741	**2.400619**	0.695058	0.026
2001	735	**2.1929**	0.737157	0.027

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
5.57	1474	**0.000**

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
1999	741	**2.400619**	0.695058	0.026
2003	700	**0.628274**	0.652386	0.025

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
49.841	1439	**0.000**

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
2001	735	**2.1929**	0.737157	0.0272
2003	700	**0.628274**	0.652386	0.0247

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
42.5	1433	**0.000**

The results of the means' tests for the second factor (nutritional perceptions) were different. The p-value of 0.000 in all cases led to the rejection of the null hypothesis at α

24 The error of accepting the null hypothesis when it is false.

= 0.05 level of significance concluding that the means of the factor scores for nutritional perceptions between the three years are different. As a result, pooling the three datasets is not appropriate and does not contribute to the purposes of this analysis.

- For factor 3 (awareness)

Group Statistics				
CODE	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
1999	733	**1.033525**	0.946415	0.035
2001	735	**0.775284**	1.011345	0.037

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
5.051	1466	**0.000**

Independent Samples Test t-test for Equality of Means		
t	**Df**	**Sig. (2-tailed)**
-12.704	1433	**0.000**

The results of the means' tests for the third factor (awareness) were identical to those for the second factor. Again, in all cases, the authors rejected the null hypothesis at α = 0.05 level of significance, concluding that the means of the factor scores for awareness are different among the three datasets and that pooling is also inappropriate.

Having rejected the null hypothesis for equal means in six out of the nine conducted tests, the authors possess sufficient evidence that the means of the factor scores change over time, and therefore, pooling the datasets is not a valid statistical method to proceed with the analysis.

APPENDIX E. SHIPMENT DATASETS AND EXPENDITURES

Month	Total Shipment Quantities
1997.08	50,338,131
1997.09	98,145,089
1997.10	89,746,078
1997.11	50,364,607
1997.12	48,423,205
1998.01	38,546,825
1998.02	38,033,593
1998.03	32,532,966
1998.04	36,219,430
1998.05	43,051,165
1998.06	43,648,014
1998.07	43,018,384
1998.08	52,367,820
1998.09	54,653,261
1998.10	71,025,595
1998.11	51,455,206
1998.12	55,931,266
1999.01	36,467,779
1999.02	38,913,502
1999.03	45,470,789
1999.04	41,040,544
1999.05	37,819,462
1999.06	41,459,573
1999.07	45,897,792
1999.08	45,635,482
1999.09	92,975,741
1999.10	107,389,618
1999.11	76,067,903
1999.12	56,017,529
2000.01	49,061,571
2000.02	51,483,976
2000.03	46,861,908
2000.04	41,751,802
2000.05	46,198,265
2000.06	52,983,344
2000.07	46,213,894
2000.08	58,253,764
2000.09	91,769,210
2000.10	109,242,644
2000.11	75,679,436

Month	Total Shipment Quantities
2000.12	54,625,935
2001.01	49,404,885
2001.02	47,100,893
2001.03	45,400,495
2001.04	47,274,569
2001.05	56,959,125
2001.06	52,784,372
2001.07	51,325,012
2001.08	60,533,455
2001.09	100,410,311
2001.10	119,107,738
2001.11	78,566,451
2001.12	70,626,908
2002.01	68,740,168
2002.02	58,040,604
2002.03	57,758,127
2002.04	56,959,584
2002.05	54,437,289
2002.06	54,108,760
2002.07	41,833,240
2002.08	53,998,620
2002.09	112,299,662
2002.10	127,682,648
2002.11	112,063,276
2002.12	89,238,158
2003.01	81,041,429
2003.02	66,658,837
2003.03	64,548,422
2003.04	71,919,671
2003.05	69,557,705
2003.06	73,897,574
2003.07	59,466,372
2003.08	66,618,738
2003.09	98,878,739
2003.10	144,907,950
2003.11	96,759,223
2003.12	93,606,991
2004.01	77,048,793
2004.02	75,697,770
2004.03	81,987,951

	Expenditure Categories			
Year/Month	**Public Relations**	**Food Service industry**	**Advertising**	**Nutrition Research**
1999.07	0.00	98,173.70	809,515.10	95,754.62
1999.08	.	25,608.02	496,185.50	5,292.35
1999.09	167,015.00	43,109.92	623,896.50	46,378.53
1999.10	11,877.00	0.00	578,708.30	5,902.04
1999.11	19,033.00	0.00	234,995.40	20,837.45
1999.12	0.00	10,459.28	712,672.30	42,399.41
2000.01	62,934.00	44,480.97	6,254.46	244,799.20
2000.02	213,500.00	38,051.00	814,571.00	110,452.00
2000.03	87,822.00	16,952.00	589,007.00	57,030.00
2000.04	222,555.00	11,239.00	467,362.00	88,205.00
2000.05	333,097.00	8,574.00	378,421.00	7,999.00
2000.06	309,735.00	16,720.00	437,149.00	94,553.00
2000.07	480,641.00	53,874.25	121,876.70	89,050.00
2000.08	27,719.00	23,308.08	0.00	9,879.30
2000.09	0.00	6,465.66	619,161.40	20.00
2000.10	10,961.00	30,222.35	677,980.40	69,909.31
2000.11	11,974.00	19,647.58	164,351.60	7,083.40
2000.12	6,700.00	18,380.78	262,048.20	56,583.28
2001.01	13,916.00	12,058.00	398,069.00	80,173.00
2001.02	.	.	.	.
2001.03	.	.	.	.
2001.04	.	.	.	.
2001.05	.	.	.	.
2001.06	.	.	.	.
2001.07	1,026,423.00	30,555.00	196,137.00	74,467.00
2001.08	48,949.00	22,736.00	348,071.00	31,163.00
2001.09	56,373.00	6,858.00	723,435.00	6,208.00
2001.10	175,870.00	25,784.00	824,844.00	51,896.00
2001.11	126,196.00	17,608.00	119,855.00	10,181.00
2001.12	94,980.00	27,399.00	443,231.00	136,501.00
2002.01	152,126.00	206,825.00	415,264.00	17,726.00
2002.02	79,604.00	29,052.00	667,701.00	15,309.00
2002.03	183,912.00	31,404.00	333,822.00	103,028.00
2002.04	279,094.00	28,013.00	576,759.00	17,533.00
2002.05	241,425.00	16,877.00	510,329.00	175,103.00
2002.06	137,451.00	18,113.00	424,071.00	44,741.00
2002.07	1,078,427.00	21,876.00	138,573.00	174,759.00
2002.08	17,397.00	13,208.00	878,175.00	4,886.00
2002.09	22,381.00	6,428.00	685,242.00	3,648.00
2002.10	315,577.00	19,420.00	909,343.00	38,396.00
2002.11	118,246.00	21,059.00	97,409.00	90.00

	Expenditure Categories			
Year/Month	**Public Relations**	**Food Service industry**	**Advertising**	**Nutrition Research**
2002.12	130,998.00	10,569.00	381,813.00	165,243.00
2003.01	194,278.00	52,199.00	514,906.00	3,744.00
2003.02	271,050.00	44,712.00	540,549.00	51,976.00
2003.03	214,235.00	25,283.00	396,440.00	3,266.00
2003.04	392,078.00	38,070.00	480,519.00	54,293.00
2003.05	106,276.00	.	450,922.00	10,839.00
2003.06	317,673.00	7,039.00	571,724.00	51,100.00
2003.07	14,855.00	645,850.00	539,764.00	345,288.00
2003.08	123,809.00	17,990.00	598,818.00	3,222.00
2003.09	581,280.00	2,335.00	532,309.00	59,387.00
2003.10	270,643.00	2,861.00	433,896.00	15,251.00
2003.11	505,176.00	6,005.00	156,647.00	15,483.00
2003.12	.	27,792.00	861,620.00	3,250.00
2004.01	251,318.00	16,465.00	844,333.00	1,828.00
2004.02	150,244.00	10,202.00	630,433.00	10,656.00
2004.03	258,177.00	42,930.00	579,551.00	609,819.00

Note: Negative amounts extracted from the financial records of the organization are represented as missing values in the dataset.

APPENDIX F. VARIABLE EVALUATION SELECTION PROCESS

Adj. R^2 w/o variable refers to the adj R^2 of the model including only the dummy variables adjusting for seasonality.

#	Variable	Adj. R^2 w/o Variable	Adj. R^2 with Variable	Change in R^2	F-Sig for change	n	Result: Selected if F-Test for change is significant at α=0.15
1	Public Relations	0.651	0.644	0.000	0.827	50	Not selected
2	Food Service	0.640	0.633	0.001	0.765	51	Not selected
3	Advertising	0.636	0.676	0.044	0.011	52	Selected
4	Nutrition Research	0.636	0.632	0.003	0.537	52	Not selected
5	Public Relations Lag 1	0.622	0.651	0.035	0.035	49	Selected
6	Public Relations Lag 2	0.603	0.613	0.017	0.156	48	Not selected
7	Public Relations Lag 3	0.661	0.716	0.057	0.004	47	Selected
8	Public Relations Lag 4	0.568	0.567	0.008	0.359	47	Not selected
9	Public Relations Lag 5	0.637	0.647	0.016	0.156	46	Not selected
10	Public Relations Lag 6	0.642	0.671	0.035	0.036	45	Selected
11	Public Relations Lag 7	0.641	0.638	0.006	0.402	44	Not selected
12	Public Relations Lag 8	0.663	0.657	0.002	0.610	43	Not selected
13	Public Relations Lag 9	0.677	0.670	0.002	0.665	42	Not selected
14	Public Relations Lag 10	0.637	0.629	0.001	0.733	41	Not selected
15	Public Relations Lag 11	0.629	0.630	0.011	0.296	40	Not selected
16	Public Relations Lag 12	0.618	0.608	0.001	0.783	39	Not selected
16	Food Service Lag 1	0.632	0.626	0.002	0.629	50	Not selected
18	Food Service Lag 2	0.613	0.605	0.001	0.730	49	Not selected
19	Food Service Lag 3	0.612	0.667	0.058	0.006	48	Selected
20	Food Service Lag 4	0.553	0.551	0.007	0.392	47	Not selected
21	Food Service Lag 5	0.607	0.632	0.031	0.058	46	Selected
22	Food Service Lag 6	0.631	0.623	0.001	0.761	45	Not selected
23	Food Service Lag 7	0.667	0.660	0.002	0.609	44	Not selected
24	Food Service Lag 8	0.665	0.686	0.027	0.067	43	Selected
25	Food Service Lag 9	0.681	0.681	0.008	0.307	42	Not selected
26	Food Service Lag 10	0.665	0.656	0.000	0.898	41	Not selected
27	Food Service Lag 11	0.638	0.651	0.021	0.129	41	Selected
28	Food Service Lag 12	0.624	0.615	0.001	0.699	40	Not selected
29	Advertising Lag 1	0.620	0.612	0.000	0.870	51	Not selected
30	Advertising Lag 2	0.616	0.611	0.003	0.519	50	Not selected
31	Advertising Lag 3	0.609	0.616	0.015	0.177	49	Not selected
32	Advertising Lag 4	0.568	0.563	0.004	0.494	48	Not selected

#	Variable	Adj. R^2 w/o Variable	Adj. R^2 with Variable	Change in R^2	F-Sig for change	n	Result: Selected if F-Test for change is significant at α=0.15
33	Advertising Lag 5	0.631	0.623	0.000	0.820	47	Not selected
34	Advertising Lag 6	0.637	0.635	0.006	0.389	46	Not selected
35	Advertising Lag 7	0.644	0.636	0.001	0.782	45	Not selected
36	Advertising Lag 8	0.665	0.684	0.025	0.072	44	Selected
37	Advertising Lag 9	0.680	0.681	0.009	0.295	43	Not selected
38	Advertising Lag 10	0.644	0.639	0.004	0.487	42	Not selected
39	Advertising Lag 11	0.638	0.641	0.012	0.250	41	Not selected
40	Advertising Lag 12	0.624	0.643	0.027	0.097	40	Selected
41	Nutrition Research Lag 1	0.620	0.616	0.005	0.438	51	Not selected
42	Nutrition Research Lag 2	0.616	0.630	0.021	0.101	50	Selected
43	Nutrition Research Lag 3	0.609	0.628	0.025	0.077	49	Selected
44	Nutrition Research Lag 4	0.568	0.563	0.005	0.474	48	Not selected
45	Nutrition Research Lag 5	0.631	0.638	0.014	0.183	47	Not selected
46	Nutrition Research Lag 6	0.637	0.641	0.012	0.236	46	Not selected
47	Nutrition Research Lag 7	0.644	0.637	0.002	0.630	45	Not selected
48	Nutrition Research Lag 8	0.665	0.663	0.006	0.371	44	Not selected
49	Nutrition Research Lag 9	0.680	0.686	0.013	0.193	43	Not selected
50	Nutrition Research Lag 10	0.644	0.635	0.001	0.802	42	Not selected
51	Nutrition Research Lag 11	0.638	0.646	0.017	0.178	41	Not selected
52	Nutrition Research Lag 12	0.624	0.614	0.001	0.754	40	Not selected
53	Nutrition Research Lag 13	0.605	0.604	0.009	0.349	39	Not selected
54	Nutrition Research Lag 14	0.596	0.609	0.022	0.158	38	Not selected
55	Nutrition Research Lag 15	0.586	0.611	0.034	0.088	39	Selected
56	Nutrition Research Lag 16	0.529	0.538	0.021	0.214	36	Not selected
57	Nutrition Research Lag 17	0.549	0.558	0.021	0.211	35	Not selected
58	Nutrition Research Lag 18	0.577	0.563	0.000	0.897	34	Not selected
59	Nutrition Research Lag 19	0.604	0.596	0.006	0.496	33	Not selected
60	Nutrition Research Lag 20	0.666	0.654	0.001	0.829	32	Not selected
61	Nutrition Research Lag 21	0.694	0.683	0.001	0.807	31	Not selected
62	Nutrition Research Lag 22	0.717	0.707	0.001	0.802	30	Not selected
63	Nutrition Research Lag 23	0.703	0.713	0.019	0.184	29	Not selected
64	Nutrition Research Lag 24	0.708	0.696	0.001	0.803	28	Not selected

APPENDIX G. MODEL TRIALS

#	ADV	PR	FS	NR	Adj R^2	F sig	ADV sig	PR sig	FS sig	NR sig	n
1	ADV 0	PR 1	FS 3	NR 2	0.729	0.000	0.008	0.295	0.037	0.203	44
2	ADV 0	PR 1	FS 3	NR 3	0.721	0.000	0.008	0.232	0.022	0.459	44
3	ADV 0	PR 1	FS 3	NR 15	0.797	0.000	0.010	0.010	0.813	0.028	27
4	ADV 0	PR 1	FS 5	NR 2	0.669	0.000	0.045	0.691	0.311	0.114	40
5	ADV 0	PR 1	FS 5	NR 3	0.668	0.000	0.027	0.547	0.225	0.123	40
6	ADV 0	PR 1	FS 5	NR 15	0.690	0.000	0.042	0.054	0.824	0.094	25
7	ADV 0	PR 1	FS 11	NR 2	0.734	0.000	0.004	0.286	0.076	0.312	33
8	ADV 0	PR 1	FS 11	NR 3	0.740	0.000	0.010	0.659	0.068	0.272	32
9	ADV 0	PR 1	FS 11	NR 15	0.818	0.000	0.002	0.002	0.030	0.010	26
10	ADV 0	PR 3	FS 3	NR 2	0.796	0.000	0.274	0.012	0.001	0.508	43
11	ADV 0	PR 3	FS 3	NR 3	0.797	0.000	0.259	0.005	0.001	0.451	43
12	ADV 0	PR 3	FS 3	NR 15	0.743	0.000	0.974	0.141	0.035	0.441	27
13	ADV 0	PR 3	FS 5	NR 2	0.733	0.000	0.303	0.029	0.397	0.334	40
14	ADV 0	PR 3	FS 5	NR 3	0.755	0.000	0.205	0.004	0.340	0.055	40
15	ADV 0	PR 3	FS 5	NR 15	0.686	0.000	0.816	0.086	0.330	0.454	25
16	ADV 0	PR 3	FS 11	NR 2	0.730	0.000	0.029	0.561	0.067	0.314	32
17	ADV 0	PR 3	FS 11	NR 3	0.753	0.000	0.125	0.348	0.065	0.080	32
18	ADV 0	PR 3	FS 11	NR 15	0.718	0.000	0.242	0.915	0.061	0.095	24
19	ADV 0	PR 6	FS 3	NR 2	0.768	0.000	0.004	0.254	0.022	0.111	38
20	ADV 0	PR 6	FS 3	NR 3	0.752	0.000	0.005	0.171	0.012	0.424	38
21	ADV 0	PR 6	FS 3	NR 15	0.684	0.000	0.044	0.700	0.163	0.387	25
22	ADV 0	PR 6	FS 5	NR 2	0.732	0.000	0.004	0.197	0.287	0.101	38
23	ADV 0	PR 6	FS 5	NR 3	0.731	0.000	0.002	0.132	0.184	0.108	38
24	ADV 0	PR 6	FS 5	NR 15	0.619	0.001	0.053	0.594	0.298	0.578	25
25	ADV 0	PR 6	FS 11	NR 2	0.752	0.000	0.004	0.149	0.028	0.263	30
26	ADV 0	PR 6	FS 11	NR 3	0.761	0.000	0.015	0.106	0.028	0.151	30
27	ADV 0	PR 6	FS 11	NR 15	0.747	0.000	0.008	0.464	0.032	0.138	23
28	ADV 8	PR 1	FS 3	NR 2	0.723	0.000	0.110	0.183	0.116	0.156	35
29	ADV 8	PR 1	FS 3	NR 3	0.702	0.000	0.149	0.173	0.118	0.975	35
30	ADV 8	PR 1	FS 3	NR 15	0.811	0.000	0.017	0.001	0.327	0.090	23
31	ADV 8	PR 1	FS 5	NR 2	0.655	0.000	0.158	0.694	0.182	0.136	35
32	ADV 8	PR 1	FS 5	NR 3	0.629	0.000	0.255	0.596	0.151	0.581	35
33	ADV 8	PR 1	FS 5	NR 15	0.727	0.000	0.017	0.010	0.970	0.012	23
34	ADV 8	PR 1	FS 11	NR 2	0.693	0.000	0.135	0.125	0.173	0.350	31
35	ADV 8	PR 1	FS 11	NR 3	0.713	0.000	0.082	0.548	0.147	0.084	30
36	ADV 8	PR 1	FS 11	NR 15	0.805	0.000	0.015	0.001	0.039	0.002	24
37	ADV 8	PR 3	FS 3	NR 2	0.804	0.000	0.137	0.005	0.002	0.423	36
38	ADV 8	PR 3	FS 3	NR 3	0.806	0.000	0.192	0.003	0.003	0.756	37
39	ADV 8	PR 3	FS 3	NR 15	0.766	0.000	0.216	0.011	0.011	0.629	25
40	ADV 8	PR 3	FS 5	NR 2	0.756	0.000	0.166	0.011	0.340	0.319	36
41	ADV 8	PR 3	FS 5	NR 3	0.769	0.000	0.290	0.002	0.298	0.209	37
42	ADV 8	PR 3	FS 5	NR 15	0.731	0.000	0.124	0.015	0.326	0.651	25
43	ADV 8	PR 3	FS 11	NR 2	0.718	0.000	0.193	0.052	0.083	0.271	31

#	ADV	PR	FS	NR	Adj R^2	F sig	ADV sig	PR sig	FS sig	NR sig	n
44	ADV 8	PR 3	FS 11	NR 3	0.778	0.000	0.049	0.061	0.063	0.016	32
45	ADV 8	PR 3	FS 11	NR 15	0.725	0.000	0.151	0.029	0.035	0.064	24
46	ADV 8	PR 6	FS 3	NR 2	0.724	0.000	0.115	0.324	0.046	0.169	37
47	ADV 8	PR 6	FS 3	NR 3	0.717	0.000	0.117	0.195	0.104	0.660	38
48	ADV 8	PR 6	FS 3	NR 15	0.652	0.000	0.066	0.592	0.121	0.366	26
49	ADV 8	PR 6	FS 5	NR 2	0.678	0.000	0.141	0.247	0.137	0.162	37
50	ADV 8	PR 6	FS 5	NR 3	0.667	0.000	0.157	0.160	0.114	0.825	38
51	ADV 8	PR 6	FS 5	NR 15	0.614	0.000	0.054	0.382	0.088	0.391	28
52	ADV 8	PR 6	FS 11	NR 2	0.681	0.000	0.144	0.228	0.125	0.301	32
53	ADV 8	PR 6	FS 11	NR 3	0.746	0.000	0.036	0.134	0.081	0.018	33
54	ADV 8	PR 6	FS 11	NR 15	0.666	0.000	0.094	0.238	0.090	0.108	28
55	ADV 12	PR 1	FS 3	NR 2	0.662	0.000	0.178	0.341	0.182	0.435	31
56	ADV 12	PR 1	FS 3	NR 3	0.653	0.000	0.281	0.400	0.247	0.869	31
57	ADV 12	PR 1	FS 3	NR 15	0.722	0.000	0.124	0.010	0.355	0.030	25
58	ADV 12	PR 1	FS 5	NR 2	0.613	0.000	0.282	0.785	0.118	0.345	29
59	ADV 12	PR 1	FS 5	NR 3	0.596	0.000	0.395	0.778	0.119	0.934	29
60	ADV 12	PR 1	FS 5	NR 15	0.592	0.003	0.354	0.071	0.807	0.092	23
61	ADV 12	PR 1	FS 11	NR 2	0.686	0.000	0.020	0.107	0.076	0.202	32
62	ADV 12	PR 1	FS 11	NR 3	0.676	0.000	0.130	0.446	0.072	0.263	31
63	ADV 12	PR 1	FS 11	NR 15	0.728	0.000	0.170	0.008	0.062	0.047	27
64	ADV 12	PR 3	FS 3	NR 2	0.770	0.000	0.852	0.002	0.002	0.998	32
65	ADV 12	PR 3	FS 3	NR 3	0.782	0.000	0.842	0.001	0.005	0.550	33
66	ADV 12	PR 3	FS 3	NR 15	0.767	0.000	0.609	0.001	0.002	0.979	27
67	ADV 12	PR 3	FS 5	NR 2	0.690	0.000	0.752	0.027	0.353	0.555	30
68	ADV 12	PR 3	FS 5	NR 3	0.711	0.000	0.950	0.006	0.400	0.268	31
69	ADV 12	PR 3	FS 5	NR 15	0.672	0.000	0.792	0.008	0.523	0.941	25
70	ADV 12	PR 3	FS 11	NR 2	0.705	0.000	0.085	0.048	0.049	0.197	32
71	ADV 12	PR 3	FS 11	NR 3	0.743	0.000	0.329	0.025	0.031	0.042	33
72	ADV 12	PR 3	FS 11	NR 15	0.684	0.000	0.448	0.022	0.045	0.294	27
73	ADV 12	PR 6	FS 3	NR 2	0.654	0.000	0.400	0.433	0.104	0.338	30
74	ADV 12	PR 6	FS 3	NR 3	0.647	0.000	0.564	0.303	0.202	0.767	31
75	ADV 12	PR 6	FS 3	NR 15	0.568	0.002	0.475	0.388	0.154	0.934	25
76	ADV 12	PR 6	FS 5	NR 2	0.606	0.000	0.244	0.309	0.093	0.299	30
77	ADV 12	PR 6	FS 5	NR 3	0.592	0.000	0.334	0.188	0.098	0.953	31
78	ADV 12	PR 6	FS 5	NR 15	0.519	0.002	0.376	0.244	0.108	0.984	27
79	ADV 12	PR 6	FS 11	NR 2	0.695	0.000	0.047	0.103	0.032	0.128	30
80	ADV 12	PR 6	FS 11	NR 3	0.711	0.000	0.185	0.047	0.022	0.089	31
81	ADV 12	PR 6	FS 11	NR 15	0.660	0.000	0.118	0.071	0.039	0.598	28

APPENDIX H. CONTRIBUTION OF SHORT AND LONG TERM VARIABLES IN THE EXPLANATORY POWER OF THE MODEL

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	DUMMY23, DUMMY22, DUMMY21[a]	.	Enter
2	Advertizing, LAGS(PUBLIC_R,1)[a]	.	Enter
3	LAGS(FOOD_SER,11), LAGS(NUTRITIO,15)[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Total Shipments

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.755[a]	.571	.512	6561246.28	.571	9.748	3	22	.000
2	.873[b]	.761	.702	2947689.04	.191	7.997	2	20	.003
3	.932[c]	.869	.818	0103272.39	.108	7.423	2	18	.004

a. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21

b. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21, Advertizing, LAGS(PUBLIC_R,1)

c. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21, Advertizing, LAGS(PUBLIC_R,1), LAGS(FOO LAGS(NUTRITIO,15)

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.02E+15	3	2.674E+15	9.748	.000[a]
	Residual	6.03E+15	22	2.743E+14		
	Total	1.41E+16	25			
2	Regression	1.07E+16	5	2.140E+15	12.768	.000[b]
	Residual	3.35E+15	20	1.676E+14		
	Total	1.41E+16	25			
3	Regression	1.22E+16	7	1.745E+15	17.099	.000[c]
	Residual	1.84E+15	18	1.021E+14		
	Total	1.41E+16	25			

a. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21

b. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21, Advertizing, LAGS(PUBLIC_R,1)

c. Predictors: (Constant), DUMMY23, DUMMY22, DUMMY21, Advertizing, LAGS(PUBLIC_R,1), LAGS(FOOD_SER,11), LAGS(NUTRITIO,15)

d. Dependent Variable: Total Shipments

LIST OF REFERENCES

Anderson, Sweeney, Williams (2003): Modern Business Statistics with Microsoft Excel, Southwestern.

Basilevsky, A., (1994): Statistical Factor Analysis and Related Methods, John Wiley & Sons Inc.

Bollen, K., (1989): Structural Equations with Latent Variables, John Wiley & Sons.

Bollen, K. and Long, S. J., (1993): Testing Structural Equation Models, SAGE Publications.

Cureton, E. and D'Agostino, R., (1983): Factor Analysis: An Applied Approach, Lawrence Erlbaum Associates Publishers.

Field, A. (2000): Discovering Statistics using SPSS for Windows, SAGE Publications.

Hamilton, (1994): Time Series analysis, Princeton.

Hatcher, L. (1994): A Step – by – Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling, SAS Institute Inc., Cary, North Carolina.

Jackson, D. and Borgatta, E., (1981): Factor Analysis and Measurement in Sociological Research, SAGE Publications Ltd.

Joreskog, K. and Sorbom, D., (1979): Advances in Factor Analysis and Structural Equation Models.

Kelloway, K., (1998): Using LISREL for Structural Equation Modeling, A Researcher's Guide, SAGE Publications.

Kim, Jae-on, and Mueller, C. (1978): Factor Analysis Statistical Methods and Practical Issues, SAGE Publications #14.

Kim, Jae-on, and Mueller, C. (1978): Introduction to Factor Analysis SAGE Publications #13.

Kline, R., (1998): Principles and Practice of Structural Equation Modeling, The Guilford Press.

Lawley, D. N. and Maxwell, A. E. (1971): Factor Analysis as a Statistical Method, American Elsevier Publishing Company Inc..

Pedhazur, E. and Schmelkin, Liora (1991): Measurement Design and Analysis, Lawrence Erlbaum Associates, Publishers.

Pindyck, S. Robert and Rubinfeld, L. Daniel (1998): Econometric Models and Economic Forecasts 4th ed., Irwin/McGraw Hill Intl Editions.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Almond Board of California
   Modesto California

4. Professor William Gates
   Naval Postgraduate School
   Monterey, California

5. Professor Kenneth H. Doerr
   Naval Postgraduate School
   Monterey, California

6. Professor Mary Malina
   Naval Postgraduate School
   Monterey, California

7. Vassileios Drakopoulos
   Naval Postgraduate School
   Monterey, California

8. Michael Georgiadis
   Naval Postgraduate School
   Monterey, California

9. Thomas Markopoulos
   Naval Postgraduate School
   Monterey, California