“A model to predict outlet stores’ ratings using value, quality, selection, and help”

AUTHORS	Khalid M. Dubas https://orcid.org/0000-0001-5271-4484 Saeed M. Dubas David J. Hill

ARTICLE INFO	Khalid M. Dubas, Saeed M. Dubas and David J. Hill (2015). A model to predict outlet stores’ ratings using value, quality, selection, and help. *Innovative Marketing* , *11*(1), 28-41
RELEASED ON	Thursday, 12 March 2015
JOURNAL	"Innovative Marketing "
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
0	0	0

Khalid M. Dubas (USA), Saeed M. Dubas (USA), David Hill (USA)

A model to predict outlet stores' ratings using value, quality, selection, and help

Abstract

This study presents a model to explain and predict outlet stores' ratings by shoppers. For this purpose we utilize three surveys of outlet store shoppers conducted by *Consumer Reports'* National Research Center to evaluate shoppers' ratings of outlet stores based on several predictors. These three surveys were conducted in 2006, 2011, and 2014. Several correlation and regression analyses were performed on the datasets. Our findings are consistent with rising income and wealth disparity in the US as represented in the growth of upscale, luxury retailers and discount stores. We focus on the discount stores, specifically the outlet stores. This study shows a significant change in shoppers' attitudes from 2006 to 2011 and 2014 as they seek more value than selection and help in outlet stores than they did during 2006.

A good scientific study must be reproducible for it to be useful in accumulating valid and reliable scientific knowledge. The findings of a reproducible study can be properly and independently verified and/or refuted by future research. A reproducible study can also be replicated by other researchers by using new data. Research studies that are objective, public, reproducible, and replicable, help to accumulative valid and reliable scientific knowledge. This study is designed to be reproducible and replicable by future researchers so we present the sources of data and utilize an open-source software R.

Keywords: outlet stores' ratings, model building, retail dynamics, reproducible research.
JEL Classification: M31.

Introduction

Consumer expenditure is about 70% of the US GDP while the yearly retail sales account for about 27% of the GDP. In 2013, the US GDP was $16.77 trillion and the total retails sales were $4.53 trillion. In 2014, there were about 1,128,112 retail stores in the US. The US *Census of Retail Trade* showed that over half of the US retailers had annual sales of less than $1 million. Less than 15% of retail stores, selling more than $5 million per year, account for almost 75% of retail sales in the US. The retailers achieve economies of scale primarily with a corporate chain where a firm owns and operates more than one store. About half of all retail sales are by chain stores; and among different retail formats, the share of sales by the chain stores has been steadily increasing over time. Corporate chains take advantage of their size to get quantity discounts, and use computers to manage inventory costs and share promotion across chain stores. Some large chains are the channel captains in their channel systems (Perreault, Cannon, and McCarthy, 2014).

Retail dynamics

Theory development is important for organizing knowledge and proper decision making. Without well developed theories, the decision makers would be faced with conflicting explanations of phenomenon. In a path-breaking book that emphasized Retail Marketing, Rosenbloom (1981) notes the importance of retail dynamics by indicating that the "retail structure is in a process of continuous change". Some changes are sudden while others are subtle and develop over a long time period. He notes that retail managers must be alert to the changing retail structure and understand the dynamics of change. For this purpose, they must recognize important developments and trends early to appropriately plan their marketing strategies. He presents several theories and hypotheses that have been offered in the retailing literature in order to explain the changing dynamics of retail structure. These theories and hypotheses include the following:

1. *The wheel of retailing hypothesis* – New retail formats enter as low-price, low-margin, low-service, and low-status retailers who over time upgrade their prices, margins, service, and status. "Department stores, supermarkets, and mass-merchandisers went through this cycle" (Perreault, Cannon, and McCarthy, 2014).
2. *The retail accordion theory* – Retailers' domination alternates between general-line, wide-assortment retailers and narrow-line, specialized retailers.
3. *The retail life cycle theory* – Retailers go through stages like early growth, accelerated development, maturity, and decline. "Recent innovators, like the Internet merchants, are still in the market growth stage" (Perreault, Cannon, and McCarthy, 2014).
4. *The differential advantage theory* – Retailers gain competitive advantage by finding and filling niches in the market.

To be successful, retailers need relevant and up-to-date information to design the right retail marketing strategy and their retailing mix that fits their

customer' needs. The retail mix includes store location, store operations, merchandising, pricing, store image, and retail promotion. Their strategy and retail mix must adapt to the changes in the environmental forces. The most important environmental changes in the last few years have been economic, competitive, and technological. Kotler and Keller (2012) note that with the onset of the Great Recession in the US in December 2007, many retailers adopted defensive response by cutting stock levels, slowing expansion, and discounting deeply. Others managed inventory carefully, cut product lines, and carefully avoided over-promoting. They note the following longer-term trends in US retailing:

1. *New retail forms and combinations* – Some examples include bookstores with coffee shops, gas stations include food stores.
2. *Growth of inter-type competition* – For example, department stores, discount stores, catalog showrooms, carry similar merchandise and compete for the same customer dollars.
3. *Competition between store-based and non-store-based retailer* – None-store-based retailing is taking business away from store-based retailers.
4. *Growth of giant retailers* – Giant retailers enjoy advantages of superior information systems, logistics, and purchasing power that smaller retailers can not match. Some giant retailers are *category killers* that dominate one product category, for example, pet food (PETCO), home improvement (Home Depot), office supplies (Staples), toys (Toys-R-Us), electronics (Best Buy), books (Barnes & Nobles), etc. Some giant retailers like Walmart have built *super centers* that combine groceries and a very large collection of nonfood merchandise.
5. *Decline of middle-market retailers* – The present-day retail market is characterized as an hourglass or a dog-bone where growth is centered at the top with luxury products, and at the bottom with discount pricing. This also reflects the higher income and wealth disparity between the haves and the have-nots in the US as the income and wealth inequality has increased over time.
6. *Growing investment in technology* – Now almost all retailers invest in technology to improve their forecasts, inventory control, electronic order processing by their vendors, and to help their customers.
7. *Global profile of major retailers* – Retailers with strong brand names are growing globally, such as The Limited, the Gap, Walmart, Benetton, IKEA, etc.
8. *Growth of shopper marketing* – Research indicates that 70% to 80% of purchase decisions are made inside a retail store. As a result, retailers have increased their efforts to influence shoppers at the point of purchase by focusing on product displays and in-store advertising.
9. *Walk-in stores and their online counterparts* – Due to changes in technology and the competitive landscape, now most stores have their online counterparts to compete with other online stores. Dubas, Hershey, and Dubas (2015) discuss the leading US-based walk-in stores and their online counterparts.

History of outlet stores

Originally outlet stores were located near factories where the products were made. At first the outlet stores only sold merchandise to employees and later to the public. The trend changed to have outlet stores located miles away from large cities. This trend was attractive due to the lower cost of land in rural undeveloped areas. Also, these areas offered tax breaks as the outlets brought new jobs opportunities. However, recently outlets are opening in large metropolitan areas. Neiman Marcus has been opening outlet stores in Dallas, Houston, and New Orleans. *Consumer Reports'* (2014) notes that "Outlet shopping has also become more convenient. In the past, centers were built far from full-price stores in big cities to avoid competing with them. But retailers have seen that outlet stores actually complement their business, so more chains are building in or close to big cities".

Outlet stores

Outlet stores are a type of corporate chain where a retailer sells products at a discount than at its regular stores. These stores have been an important part of retailing in the US as they offer quality products at a discount over full retail prices at chain stores elsewhere. The US shoppers were severely affected by the Great Recession (December 2007 to June 2009) as many homeowners lost their jobs and their houses and had to cut back on purchases. The income and wealth disparity has been increasing in the US over the last several years and this has reflected in the growth of retail sales at the high-end, luxury stores, and at discount stores, while the mid-range retail stores have experienced declines. *Consumer Reports'* (2014) notes that "Outlet malls certainly are a big business – and getting bigger... From 2006 through 2012, the amount American shoppers spent at outlet centers grew 41 percent, vs. 9 percent for traditional malls. By the end of this year, 50 new outlet center will have been built since 2006; only three new retail malls have opened in the past eight years. The industry estimates that

consumers will spend $42 billion in outlet stores this year, up from $24.3 billion in 2012".

Characteristics of outlet stores

These are some of the main characteristics of outlet stores.

1. *Location.* Most are about 45 minutes' of drive from large cities.
2. *Atmosphere.* Outlets usually have tile floors with no music playing.
3. *Prices.* Made-for-the-outlet goods are usually designed so that they can be sold at a lower price since their full-price versions are better either due to better design, materials, appearance, or longevity (*Consumer Reports*, 2014). *Consumer Reports'* (2014) representatives found that most of the goods that they purchased were 3 to 72 percent less expensive at the outlets than similar items they bought at retail. Most of their readers were happy with the outlet deals as 64 percent of store visits yielded great value, and at 34 percent of stores, the prices were much lower than the sale prices at regular stores... However 17 percent of the readers said that the prices were higher than they expected" (*Consumer Reports*, 2014). So some prices can be misleading and some reference prices may be inflated.
4. *Merchandise quality.* Outlet stores have been known to sell irregular stock with minor flaws, off season, or older and discontinued items. However, most goods are now manufactured overseas, so damaged items are removed before they are shipped to the US. Also, retailers have become better at forecasting demand so there are fewer production runs (*Consumer Reports*, 2014). Outlet stores normally sell goods in unlimited quantities to consumers.
5. *Product selection. Consumer Reports'* (2014) notes that 38 percent of their survey respondents said that selection at outlet stores was as good as at regular stores, and 25 percent said that there was a wider selection of goods. This observation was confirmed by the *Consumer Reports* in their own trips to outlet stores for some categories of products.
6. *Help at outlet stores.* Service received good scores, in *Consumer Reports'* (2014) survey with just 14 percent of respondents calling outlet service fair, poor, or very poor, down from 20 percent in 2010.
7. *Return policies.* Return policies vary at outlet stores. At some outlet stores sales are final and items cannot be returned. Even if an outlet accepts returns, the drive back to the outlet will cost the customer. This curtails the volume of returns.

Characteristics of shoppers at outlet stores:

These are the main characteristics of shoppers at outlet store as described in the literature.

1. Most shoppers are tempted to spend more money because of lower cost. One study found that visitors spent 79% more money per visit at outlet centers than ordinary malls. Big discounts of 60% off are common than at outlet malls. Experts say that 82 percent of products are made specifically for the outlets (Yager November 2, 2010).
2. Shoppers seek coupon books to save money. Consumers ask about AARP (the American Association of Retired Persons), student, and military discounts.
3. Shoppers want a better value at a lower price.
4. Consumers are loyal to outlet stores regardless of changes in chain or market mix. (Lin and Gijsbrechts June, 2014).
5. Blanding (2014) discusses research by Donald Ngwe who compared customers that shopped at outlets and at retail stores. Ngwe found both groups were identical in terms of demographics, including income and zip codes. However, they differed on two variables: their willingness to travel and the degree to which they cared about quality.
6. People think they are getting a better deal on their purchases at outlet stores.
7. Outlet mall shopping trips with family and friends can become enjoyable social events.
8. Shoppers are pleased overall with outlet stores. Sixty percent of shoppers think that outlets offer an exceptional value. People who are willing to travel to outlet stores care the least about quality. Retailers that are opening outlet stores say the risk is minimal since most customers do not crossover (Ngwe, 2011).

Understanding customer preferences is important for retail store managers so they could design their retail marketing programs to better serve the needs of their customers. A useful method for understanding customer preferences is by conducting a survey of store shoppers. The *Consumer Reports* regularly contacts its subscribers to conduct its Outlet Store Shopper Survey through its National Research Center. This study evaluates three surveys conducted by *Consumer Reports* in 2006, 2011, and 2014 to assess shoppers' ratings of outlet stores. Here, we develop a general model to explain and predict shoppers' perceived ratings of outlet stores based on several predictors. This general model is specified and tested in two forumulations: (1) an additive effects model, and (2) a multiplicative effects model.

Research questions

The following research questions are explored in this study.

1. *Research Question #1. What factors are associated with Customers' Ratings of Outlet Stores?*
2. *Research Question #2. What are the relationships among Customers' Ratings of Outlet Stores, and their ratings of Value, Quality, Selection, and Help at outlet stores for the three surveys?*
3. *Research Question #3. Which model adequately explains Customers' Ratings of Outlet Stores for each survey?*
4. *Research Question #4: Which model adequately explains Customer' Ratings of Outlet Stores over time?*

Reproducible and replicable research

Scientific research should be valid, reliable, public, reproducible, replicable, refutable, and cumulative. A lot of published research in the Social Sciences, however, can not be reproduced by the reader due to the unavailability of the data, or software, or code that was used to produce the published research findings. An inability of others to reproduce a published study limits independent verification of its validity and it limits their attempts to properly replicate the study with new data thus limiting the reliability of the published study. The social sciences, including business and economics, lag behind the natural and physical sciences in this regard.

This study is designed to be reproducible, refutable, and replicable by the reader so we list here the tools/software that we have utilized for this study. We have utilized a Markdown template in RStudio GUI editor, the code was written in R software (version 3.1.1; 2014-07-10), and numerous R packages were utilized including *MASS*, *psych*, *car*, *kniter*, and *stargazer*. *MikTeX* (*LaTeX* for Windows), *knitr*, and *pandoc* were utilized to convert Markdown source document into a pdf output file that is the report of this study. RStudio automates many of these steps in producing the output file. The Markdown source document included the study narrative along with the R code. The output document included the narrative, tables, and graphs. All of these technologies/software are open source and freely available to the reader. We gratefully acknowledge the creators and supporters of these technologies/software that work on multiple platform like Windows, Mac OS X, Linux, etc. For this study, we used a Lenovo laptop Twist computer that has Windows 8.1 pro with Media Center (64 bit OS), Intel core i5, quad-core, 1.7MHz CPU, and 4GB of RAM.

Description of survey datasets

The *Consumer Reports* conducts a survey of its subscribers to obtain Customers' Ratings of Outlet Stores (Reader.Score) and also their ratings of Value, Quality, Selection, and Help at these outlet stores. The definitions and research methodology for these surveys are described in the respective *Consumer Reports* publications and also presented below. *Consumer Reports* notes that most outlet stores bear the names of the manufacturer.

List of variables. These are the variables for which *Consumer Reports* (2006, 2011, and 2014) collected data from its subscribers. These variables measured the experiences of the *Consumer Reports'* readers and not necessarily those of the general population.

1. *Reader Score* (Customers' Ratings of Outlet Stores) – "How respondents rated their overall satisfaction with their shopping experience and may be based on factors in addition to those listed in the survey results".
2. *Value* – Reflects readers' judgments, considering price and quality (*Consumer Reports*, 2006).
3. *Quality* – Quality of merchandise.
4. *Selection* – Selection of merchandise.
5. *Help* – Staff helpfulness.

The Reader Score was measured on a scale of 0 to 100. *Consumer Reports* (2014) notes that "A score of 100 means all respondents were completely satisfied; 80 would mean very satisfied, on average; and 60, fairly well satisfied". Respondents also rated the outlet stores on the Value they got for their money, the Selection and Quality of merchandise, and the store's Help on six-point scales ranging from "very poor" to "excellent". Ratings represent mean scores for each category, converted to a five-point better-to-worse scale.

Methodology of customer surveys for 2014, 2011, and 2006. There were two missing values for Help variable for 2014 and also for 2011 datasets. Both missing values for Help in each survey were replaced by the mean value for Help in each survey. Table 1 presents metadata for the three surveys under study here.

Table 1. Consumer reports surveys of outlet stores, 2006, 2011 and 2014

Month & year	# Stores	Response and predictor variables	# Subscribers	# Store visits
December 2014	53	Reader.score ~ value, quality, selection, and help	15,789	25,441
November 2011	58	Reader.score ~ value, quality, selection, and help	17,753	38,846
May 2006	33	Reader.score ~ value, quality, selection, and help	6,038	11,300

All tables (except Table 1) in this study were produced by the *stargazer* package in R (Hlavac, 2014).

The survey datasets and summary statistics

Research Question #1. What factors are associated with Customers' Ratings of Outlet Stores?

To address Research Question #1, we present here, the *Consumer Reports'* three survey datasets and their summaries. These surveys indicate the variables that are relevant in explaining and predicting outlet stores' ratings by customers. The 2014 survey information is given in Tables 2 and 3; the 2011 survey information is given in Tables 4 and 5; and the 2006 survey information is given in Tables 6 and 7 (see in Appendix).

Relationships among variables

Research Question #2. What are the relationships among Customers' Ratings of Outlet Stores, and their ratings of Value, Quality, Selection, and Help at outlet stores for the three surveys?

To answer Research Questions #2, we evaluate the correlation matrices for the variables under study in the three *Consumer Reports'* surveys.

Correlation matrices

The Pearson product-moment correlation coefficients were obtained by using the *psych* package in R. Tables 8, 9, and 10 show these correlation coefficients and their *p*-values, respectively, for 2014, 2011, and 2006 (see in Appendix). These correlation coefficients show approximately similar results over time. Value, Quality, Selection, and Help have highly significant positive correlations with Reader Score (response variable). In addition, Quality has a highly significant positive correlation with Value, Selection, and Help. All correlations are positive and significantly different from zero, except the correlation between Help and Value in 2011 which is positive but insignificant.

The correlation coefficient between the response variable and any predictor variable, however, does not exclude the influence of other predictor variables. In order to show the independent influence of each predictor variable on the response variable, therefore, we need to utilize multiple regression analysis that will include all predictor variables simultaneously in the model, thus measuring the independent influence of each predictor on the response variable.

Next, we specify the general model of the influence of the predictor variables on the response variable and present two different specifications of this model for estimation using multiple regression analyses.

General model and its specifications

This study proposes the following general model to describe Customers' Ratings of Outlet Stores based on their perceptions of store Value, Quality, Selection, and Help:

Equation #1A: $StoreEvaluation = f(Value, Quality, Selection, Help)$

Here, Store Evaluation is measured by the Reader Score that was provided by the readers/subscribers of *Consumer Reports* who participated in its surveys to evaluate outlet stores.

We assume an additive and a multiplicative specification for this general model.

Model specification #1: each predictor variable has an additive effect on the response variable

A linear and additive relationship among the predictor variables is specified as follows:

Equation #2A: $StoreEvaluation = \beta_0 + \beta_1 Value + \beta_2 Quality + \beta_3 Selection + \beta_4 Help + \varepsilon.$

Model specification #2: each predictor variable has a multiplicative effect on the response variable

This model specifies a multiplicative effect of each predictor variable. This non-linear model is converted into a linear and additive model by taking natural logarithms on both sides of the equation as follows:

Equation #3A: $log(StoreEvaluation) = \gamma_0 + \gamma_1 log(Value) + \gamma_2 log(Quality) + \gamma_3 log(Selection) + \gamma_4 log(Help) + \psi.$

Both model specifications (Equation #2A and #3A) are linear in parameters that can be estimated by linear regressions.

Multiple regression analyses and results

We evaluated four types of regression models for each survey dataset. These models are (1) multiple linear regression, (2) robust linear regression (using MM method), (3) log-log linear multiple regression model, and (4) robust log-log linear regression model (using MM method). Fox (2008) describes robust linear regression using "M" method and "MM" method. Adler (2012) notes that robust regression methods can be helpful if there are heteroscedasticity and outlier problems. The rlm function in the *MASS* package utilizes MM-estimation to fit a model to the data. We use Bisquare robust regression using method = MM as described by Fox and Weisberg (2013). The robust regressions reduce the influence of extreme values on the regression coefficients.

Research Question #3. Which model adequately explains Customers' Ratings of Outlet Stores for each survey?

To answer Research Question #3, we run multiple regression analyses for each survey separately. Initially, we included Value, Quality, Selection, and Help to predict Reader Score. However, for every regression equation, the coefficient for Quality was positive and insignificant. This is most likely due to multicollinearity in terms of significant positive correlations between Quality and Value and between Quality and Selection for each of the three surveys. Therefore, we decided to exclude Quality from all regression analyses. This led to a slight decrease in R-squared values for each model. The revised models are given below:

Equation #1B: $StoreEvaluation = f(Value, Selection, Help)$

Equation #2B: $StoreEvaluation = \beta_0 + \beta_1 Value + \beta_2 Selection + \beta_3 Help + \varepsilon.$

Equation #3B: $log(StoreEvaluation) = \gamma_0 + \gamma_1 log(Value) + \gamma_2 log(Selection) + \gamma_3 log(Help) + \psi.$

The regression results are reported and discussed below.

2014 Consumer Reports survey

The overall conclusions are consistent for all four types of regression models for the 2014 dataset in Table 11. All four types of regression equations have positive coefficients and they are consistent in terms of their relative sizes such that Value is the most important predictor, followed by Selection, and Help. The intercepts are positive and significant in all four regression analyses indicating the mean effects of the excluded variables. The OLS (log-log) regression model explains about 76% (75%) of the variability in its response variable. The standard errors for these regression coefficients are small so these are reliable estimates for explanation and prediction.

The log-log regression model (using natural logarithm) estimates the elasticities of Value, Selection, and Help, as they influence the response variable. These regression coefficients indicate that the Store Ratings go up by 11 percentage points as Value increases by 1 percent, the Store Ratings go up by 6 percentage points as Selection increases by 1 percent; the Store Ratings go up by 3.7 percentage points as Help increases by 1 percent. For the elasticity interpretation of the regression coefficients of a log-log regression model, the reader is referred to Hatekar (2010).

Tests for heteroskedasticity for 2014 dataset. We performed Breusch-Pagan-Godfrey (studentized B-P test) and Goldfeld-Quandt tests (delete 1/3 of middle obs) to estimate the OLS regression equation on the first 1/3 and the last 1/3 of the data. The p-values were not less than 0.05 so the null hypothesis of homosekdasticity could not be rejected for the OLS regression model. Similarly, the log-log model was tested for the presence of homoskedasticity and the null hypothesis could not be rejected in the above two tests. These tests were not performed for the robust regressions for the 2014 dataset.

2011 Consumer Reports survey

The results for regression analyses for 2011 data set are presented in Table 12. They show that all four types of regressions have positive coefficients and they are consistent in terms of relative impact of variables such that Value is most important, followed by Selection, and Help. The intercepts are positive and significant in all four regression analyses indicating the mean effects of the excluded variables. The OLS (log-log) regression model explains about 80% (78%) of the variability in its response variable. The standard errors for regression coefficients are small so these are reliable estimates for explanation and prediction. The robust regressions reduce the influence of extreme values on the regression coefficients.

Tests for heteroskedasticity for 2011 dataset. We performed Breusch-Pagan-Godfrey (studentized B-P test) and Goldfeld-Quandt tests (delete 1/3 of middle obs) to estimate the OLS regression equation on the first 1/3 and the last 1/3 of the data. The *p*-values were not less than 0.05 so the null hypothesis of homoskedasticity could not be rejected for the OLS regression model. Similarly, the log-log model was tested for the presence of heteroskedasticity and the null hypothesis could not be rejected in the above two tests. These tests were not performed for the robust regressions for the 2011 dataset.

2006 Consumer Reports survey

The results for the regression analyses for the 2006 dataset are given in Table 13. They show that all regression coefficients are positive and significant as in the case of 2011 and 2014 datasets. However, unlike 2011 and 2014 datasets, here Selection is the most important variable for all four types of regressions. Help is the second most important variable while Value is the least important of the three predictors. The intercepts are positive and significant in all four regression analyses indicating the mean effects of the excluded variables. The OLS(log-log) regression model explains about 80%(79%) of the variability in the response variable. The standard errors for regression coefficients are small so these are reliable estimates for explanation and prediction.

Tests for heteroskedasticity for 2006 dataset. We performed Breusch-Pagan-Godfrey (studentized B-P test) and Goldfeld-Quandt tests (delete 1/3 of middle obs) to estimate the OLS regression equation on the first 1/3 and the last 1/3 of the data. The *p*-values were not less than 0.05 so the null hypothesis of homosekdasticity could not be rejected for the OLS regression model. Similarly, the log-log model was tested for the presence of heteroskedasticiy and the null hypothesis could not be rejected in the above two tests. These tests were not performed for the robust regressions for the 2006 dataset.

The robust linear and the robust log-log regression models provide the best explanation of Customers' Ratings of Outlet Stores for each survey.

Regression analyses across the three surveys

Research Question #4: Which model adequately explains customer' Ratings of Outlet Stores over time?

To answer Research Question #4, OLS regressions for the three surveys are presented in Table 14 and the log-log regressions are presented in Table 15 to facilitate comparisons across years. The robust linear and the robust log-log regressions are presented in Tables 16 and 17 respectively. The results of the 2014 and 2011 surveys are similar to each other but they are different from those of the 2006 survey. It is apparent that the customer preferences have shifted over time since 2006 as Value has become the most important predictor followed by Selection and Help for respondents who shopped at outlet stores. The robust OLS regressions and the robust log-log regressions provide the best explanations of Customer' Ratings of Outlet Stores over time.

Conclusion

This study developed a model to explain and predict Customers' Ratings of Outlet Stores based on their perceptions of Value, Quality, Selection, and Help provided by these stores. Both linear and non-linear models were specified and tested using multiple regressions. We noticed that Value was less important than Selection and Help in 2006 but over time, Value became the most important predictor of outlet store ratings as also noted by *Consumer Reports* (2014). This is consistent with increasing income and wealth disparity in the US and increasing sales at upscale as well as at discount stores – the hourglass or dog-bone effect noted by Kotler and Keller (2012). Other important variables that determine store ratings are Selection and Help.

Quality is highly correlated with Value and other predictors so it was dropped from our regression models to avoid multicollinearity. The revised model and its various specifications are given below:

Equation #1B: $StoreEvaluation = f(Value, Selection, Help)$

Here all the predictor variables have a positive influence on the response variable.

The linear and additive model in Equation #2B provides a good estimation of the relationship among the response variable and the predictor variables for all three surveys:

$$StoreEvaluation = \beta_0 + \beta_1 Value + \beta_2 Selection + \beta_3 Help + \varepsilon.$$

The linear and additive model of the log-log relationships in Equation #3B also provides a good estimation of the relationships among the response variable and the predictor variables in the three surveys. An attractive feature of the log-log models is that their regression coefficients represent the elasticities of the response variable with respect to each predictor variable.

Equation #3B: $log(StoreEvaluation) = \gamma_0 + \gamma_1 log(Value) + \gamma_2 log(Selection) + \gamma_3 log(Help) + \psi.$

Overall, the best models for each year are the robust linear regression models given in Table 16 and the robust log-log models given in Table 17. The results of various regression models for the years 2011 and 2014 are remarkably similar but they differ from those of the 2006 survey. Apparently, the Great Recession, during 2008-2009 in the US, altered shoppers' attitudes towards shopping by making Value the most important determinant of Outlet Store Ratings.

The robust linear regression model for the 2006 and 2014 surveys are given below:

2006 survey: $StoreEvaluation = 58.9584 + 0.9091 Value + 2.8154 Selection + 2.015 Help$

2014 survey: $60.2785 + 3.0418 Value + 1.277 Selection + 1.2552 Help$

The robust log-log linear regression models for the 2006 and 2014 surveys are given below:

2006 survey: $log(StoreEvaluation) = 4.123 + 0.0446 log(Value) + 0.103 log(Selection) + 0.0466 log(Help)$

2014 survey: $log(StoreEvaluation) = 4.1175 + 0.1135 log(Value) + 0.0575 log(Selection) + 0.0384 log(Help)$

Good scientific research should be valid and reliable. The validity of a reproducible study can be verified independently by other researchers. The reliability of a published study can be established by other researchers who replicate a published study by using different data. So, without reproducibility, the findings of a study may have

limited validity and/or reliability. Therefore, to be reproducible and replicable, this study has presented the sources of survey datasets that are utilized here as well as the survey datasets in raw and summary formats. We have also indicated the hardware and all of the software technologies that we utilized here so the reader could reproduce and replicate this study. Independently verifiable studies are required for building valid and reliable cumulative knowledge.

References

1. Adler, Josehp (2012). *R in a Nutshell: A Desktop Quick Reference.* Sebastopol, CA: O'Reilly Media.
2. Blanding, Michael (September 29, 2014). Why do outlet stores exist? Research & Ideas. Retrieved from http://hbswk.hbs.edu/item/7638.html.
3. Consumer Reports (September 1992). The Off-Price Option; Looking for a better deal, *Consumer Reports*, 57(9), p. 559.
4. Consumer Reports (May 2006). Outlet Stores: Where to Shop & How to Save Big Bucks, *Consumer Reports*, pp. 19-25.
5. Consumer Reports (November 2011). Outlet Stores: Guide to the Survey, *Consumer Reports*, pp. 1-3.
6. Consumer Reports (December, 2014). The Truth About Outlet Malls; Not Everything is a bargain, but if you go with eyes wide open, you will find some good deals, *Consumer Reports*, pp. 16-18.
7. Dubas, Khalid M., Lewis Hershey and Saeed M. Dubas (2015). An Evaluation of the Walk-in and Online Counterparts of the Leading US Stores, *Academy of Marketing Studies Journal*, in press.
8. Fox, John and Sanford Weisberg (2011). *An R Companion to Applied Regression.* Thousand Oaks, CA: Sage Publications, Inc.
9. Fox, John and Sanford Weisberg (October 8, 2013). Robust Regression. Retrieved from: http: //users.stat.umn.edu/~sandy/courses/8053/handouts/robust.pdf.
10. Fox, John (2008). *Applied Regression Analysis and Generalized Linear Models.* Thousand Oaks, CA: Sage Publications, Inc.
11. Hatekar, Neeraj R. (2010). *Principles of Econometrics: An Introduction [Using R*]. Thousand Oaks, CA: Sage Publications, Inc.
12. Hlavac, Marek (2014). Stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables. R package version 5.1. Available at: http://CRAN.R-project.org/package=stargazer.
13. Kleiber, Christian and Achim Zeileis (2008). *Applied Econometrics with R*, New York, NY: Springer-Science+Business Media LLC.
14. Lin, Arjen Van and Els Gijsbrechts (2014). Shopper Loyalty to Whom? Chain Verses Outlet Loyalty in the Context of Store Acquisitions, *American Marketing Association*, 51 (3), pp. 352-370.
15. Rosenbloom, Bert (1981). *Retail Marketing*, New York: NY: Random House, Inc.
16. Kotler, Philip and Kevin Keller (2012). *Marketing Management.* Upper Saddle River, NJ: Prentice Hall.
17. Perreault, William, Joseph Cannon, and E.J. McCarthy (2014). *Basic Marketing: A Marketing Strategy Planning Approach*, New York, NY: McGraw-HJill/Irwin Inc.
18. Xie, Yihui (2014). *Dynamic Documents with R and knitr.* Boca Raton, FL: Taylor & Francis Group.
19. Yager, Fred (November 2, 2010). Ever Wonder How Those Factory Outlets Work? *Consumer* Affairs. Retrieved from: http://www.consumeraffairs.com/news04/2010/11/ever-wonder-how-those-factory-outlet-work-and-if-you-are-really-saving-any-money.html.

Appendices

Table 2. Ratings of outlet stores in 2014 survey dataset

	Reader.Score	Value	Quality	Selection	Help
Bon.Worth	83	4	4	4	4
L.L.Bean	83	4	5	4	4
Haggar	82	4	4	4	4
Carters	82	3	4	4	3
OshKosh.Bgosh	81	4	4	4	4
Jockey	81	3	4	4	4
Bose	81	3	5	5	5
Coach.Coach.Mens	81	3	4	3	4
Vitamin.World	81	3	4	4	4
Bath.n.Body.Works	81	3	4	4	4
CorningWare.Corelle	80	3	4	4	3
Harry.n.David	80	3	4	5	4
BrooksBrothers	80	3	4	3	4
Chicos	80	3	4	3	4
Izod	79	4	4	3	3
Wilsons.Leather	79	3	4	3	3

Table 2 (cont.). Ratings of outlet stores in 2014 survey dataset

	Reader.Score	Value	Quality	Selection	Help
Van.Heusen	79	3	3	3	3
Kitchen.Collection	79	3	3	4	3
Easy.Spirit	78	3	4	3	3
Leggs.Hanes.Bali	78	3	4	3	3
Columbia.Sportswear	78	3	4	3	3
Bass	78	3	4	3	3
Reebok	78	3	4	3	3
VF.Outlet	78	3	3	3	5
Dressbarn	78	3	3	3	3
Crocs	78	3	4	4	3.137
Skechers.USA	77	3	4	4	3
Talbots	77	3	4	3	3
Coldwater.Creek	77	3	3	3	4
Lands.End	77	3	4	3	3
Clarks.Bostonian	77	3	4	3	3
Jones.New.York	77	3	3	2	4
Eddie.Bauer	77	3	4	3	3
Loft	77	3	3	3	3. 137
Ann.Taylor	77	3	3	3	3
The.Childerns.Place	76	3	3	3	2
Aeropostale	76	3	3	3	3
Rockport	75	3	4	3	3
Tommy.Hilfiger	75	3	4	3	3
Banana.Republic	75	3	3	3	2
Nautica	75	3	4	3	3
Famous.Footwear	74	3	3	3	3
Adidas	74	3	4	3	2
Polo.Ralph.Lauren	73	2	3	3	3
Pottery.Barn	73	2	3	3	2
Lane.Bryant	73	2	2	2	3
J.Crew	73	3	3	3	3
Under.Armour	73	2	4	3	3
American.Eagle	72	2	3	3	3
Old.Navy	72	2	2	3	1
Gap	72	2	2	2	1
Kike	72	2	3	3	2
Levis	72	2	3	3	2

Table 3. Summary of outlet stores in 2014 Survey dataset

Statistic	N	Mean	St. dev.	Min	Max
Reader.Score	53	77.245	3.150	72	83
Value	53	2.925	0.513	2	4
Quality	53	3.585	0.663	2	5
Selection	53	3.245	0.617	2	5
Help	53	3.137	0.809	1.000	5.000

Table 4. Ratings of outlet stores in 2011 Survey dataset

	Reader.Score	Value	Quality	Selection	Help
Jockey	81	4	4	3	4
Carters	80	4	4	4	3
Harry.n.David	79	3	4	4	5
Corningware	79	4	4	3	3
Izod	79	4	4	3	3
Van.Heusen	79	4	3	4	4
Coach	79	4	5	3	4
OshKosh.Bgosh	78	4	4	3	3

Table 4 (cont.). Summary: outlet stores data: 2011

	Reader.Score	Value	Quality	Selection	Help
L.L.Bean	78	4	4	3	4
Leggs.Hanes.Bali	77	4	3	3	3
Childrens.Place	77	4	3	3	2
Easy.Spirit	77	4	4	2	3
Totes.Isotner	77	4	3	3	3
Haggar	77	4	3	3	3
VF.Outlet	77	4	3	3	1
Le.Gourmet.Chef	77	3	3	4	3
Kitchen.Collection	76	3	3	3	3
Vitamin.World	76	3	3	3	4
Brooks.Brothers	76	4	4	3	4
G.H.Bass	76	4	3	3	2
Bostonian.Clarks	76	4	4	2	3
Stride.Rite	76	4	4	2	3
Lands.End.Inlets	75	4	4	3	3
Rockport	75	4	4	2	3
Wilson.Leather	75	4	3	3	3
Reebock	75	4	3	2	2
Black.n.Decker	75	4	3	3	3
Dress.Barn	75	3	3	3	4
Tommy.Hilfiger	75	3	4	3	3
Eddie.Bauer	75	4	3	2	3
Bose	74	3	5	4	4
Aeropostate	74	3	2	3	3
Nautica	74	3	4	2	3
Timberland	74	3	3	2	2
Sketchers	74	3	3	3	3
Fossil	74	3	3	3	3
Ann.Taylor	73	3	3	2	3
Naturalizer	73	3	3	2	3
Samsonite	73	3	3	3	4
Kenneth.Cole	73	3	3	2	3
Gymboree	73	3	4	3	3
Jones.New.York	72	3	4	2	3
Factory.Brand.Shoes	71	3	3	2	2
Nine.West	71	3	3	2	2
Calvin.Klein	71	3	3	2	2
Banana.Republic	71	3	3	2	1
PacSun	71	2	2	2	2
Pottery.Barn	71	3	3	3	2
Polo.Ralph.Lauren	70	3	4	2	2
J.Crew	70	3	3	2	2
Adidas	70	3	3	2	1
Levis	70	3	3	2	2
Sunglass.Hut	70	2	4	3	3
Gap	70	3	2	2	1
Dockers	69	3	3	2	2
Lane.Bryant	69	3	2	2	3
Nike	68	3	3	2	1
Casual.Male.XL	67	2	2	1	3

Table 5. Summary of outlet stores in 2011 Survey dataset

Statistic	N	Mean	St. dev.	Min	Max
Reader.Score	58	74.259	3.317	67	81
Value	58	3.362	0.583	2	4
Quality	58	3.310	0.681	2	5
Selection	58	2.603	0.647	1	4
Help	58	2.793	0.874	1	5

Table 6. Ratings of outlet stores in 2006 Survey dataset

	Reader.Score	Value	Quality	Selection	Help
L.L.Bean	81	4	4	3	4
Lenox	81	5	4	4	4
Jockey	81	3	3	3	4
Pfaltzgraff	81	3	3	3	4
Mikasa	81	4	4	4	4
Carters	80	4	3	4	3
OshKosh.Bgosh	80	4	4	4	3
Harry.n.David	79	3	4	5	5
Corning	79	3	3	4	3
Kitchen.Collection	79	3	3	4	3
Van.Heusen	78	3	3	3	3
Lands.End	78	3	3	3	4
Leggs.Hanes.Playtex	78	3	3	3	3
Bass	77	3	3	3	3
Coach	77	3	4	3	4
Greg.Norman.Reebok	77	3	3	3	3
VF.Outlet	76	4	3	3	2
Easy.Spirit	76	3	3	3	3
Eddie.Bauer	76	3	3	3	3
Black.n.Decker	76	3	3	3	3
Dress.Barn.Westport	74	2	2	3	3
Banna.Republic	73	2	2	3	2
Liz.Claiborne	73	2	3	3	2
Old.Navy	72	2	1	3	1
Pottery.Barn	72	2	2	3	2.935
Brooks.Brothers	72	3	3	2	3
Factory.Brand.Shoes	72	2	3	2	3
Gap	71	2	2	2	2
Saks.Off.5th	71	3	3	2	2.935
Toy.Liquidators	71	1	2	2	2
Polo.Ralph.Lauren	70	2	3	3	2
Tommy.Hilfiger	70	3	3	2	2
Nike	69	2	2	2	1

Table 7. Summary of outlet stores in 2006 Survey dataset

Statistic	N	Mean	St. Dev.	Min	Max
Reader.Score	33	75.788	3.871	69	81
Value	33	2.879	0.820	1	5
Quality	33	3.939	0.704	1	4
Selection	33	3.030	0.728	2	5
Help	33	2.935	0.899	1.000	5.000

Table 8. Correlation matrix for 2014 Dataset: *r* (lower triangular) and *p*-value (upper-triangular)

	Reader.Score	Value	Quality	Selection	Help
Reader.Score	0	0	0.00000	0.00000	0
Value	0.773	0	0.00000	0.002	0.00004
Quality	0.657	0.584	0	0.00000	0.00004
Selection	0.631	0.424	0.582	0	0.0003
Help	0.712	0.535	0.533	0.476	0

Table 9. Correlation matrix for 2011 dataset: *r* (lower triangular) and *p*-value (upper-triangular)

	Reader.Score	Value	Quality	Selection	Help
Reader.Score	0	0	0.0002	0.00000	0.0001
Value	0.740	0	0.004	0.025	0.389
Quality	0.477	0.375	0	0.005	0.0002
Selection	0.653	0.294	0.364	0	0.0001
Help	0.503	0.115	0.464	0.504	0

Table 10. Correlation matrix for 2006 dataset: *r* (lower triangular) and *p*-value (upper-triangular)

	Reader.Score	Value	Quality	Selection	Help
Reader.Score	0	0.00000	0.00004	0.00000	0.00000
Value	0.740	0	0.00000	0.002	0.001
Quality	0.648	0.744	0	0.004	0.00000
Selection	0.734	0.530	0.491	0	0.001
Help	0.755	0.543	0.737	0.531	0

Table 11. OLS, Robust linear, OLS log-linear, and Robust log-linear regression models: 2014 dataset

	Dependent variable			
	Outlet store ratings		Log (Outlet store ratings)	
	OLS	*Robust linear*	*OLS*	*Robust linear*
	(1)	(2)	(3)	(4)
Value	2.975*** (0.503)	3.042*** (0.517)		
Selection	1.388*** (0.402)	1.277*** (0.413)		
Help	1.259*** (0.329)	1. 255*** (0.337)		
Log (Value)			0.110*** (0.019)	0.114*** (0.020)
Log (Selection)			0.061*** (0.018)	0.058*** (0.018)
Log (Help)			0.037*** (0.012)	0.038*** (0.012)
Constant	60.090*** (1.399)	60.280*** (1.436)	4.118*** (0.020)	4.118*** (0.021)
Observations	53	53	53	53
R^2	0.777		0.759	
Adjusted R^2	0.763		0.745	
Residual std. error (df = 49)	1.533	1.575	0.021	0.020
F statistic (df = 3; 49)	56.830***		51.550***	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 12. OLS, Robust linear, OLS log-linear, and Robust log-linear regression models: 2011 dataset

	Dependent variable			
	Outlet store ratings		Log (Outlet store ratings)	
	OLS	*Robust linear*	*OLS*	*Robust linear*
	(1)	(2)	(3)	(4)
Value	3.461*** (0.358)	3.558*** (0.378)		
Selection	1.760*** (0.371)	1.791*** (0.391)		
Help	0.987*** (0.264)	1.032*** (0.279)		
Log (Value)			0.139*** (0.016)	0.145*** (0.017)
Log (Selection)			0.060*** (0.013)	0.058*** (0.013)
Log (Help)			0.033*** (0.008)	0.035*** (0.009)
Constant	55.280*** (1.299)	54.770*** (1.372)	4.052*** (0.019)	4.046*** (0.020)

Table 12 (cont.). OLS, Robust linear, OLS log-linear, and Robust log-linear regression models: 2011 dataset

	Dependent variable			
	Outlet store ratings		Log (Outlet store ratings)	
	OLS	*Robust linear*	*OLS*	*Robust linear*
	(1)	(2)	(3)	(4)
Observations	58	58	58	58
R^2	0.805		0.787	
Adjusted R^2	0.794		0.775	
Residual std. error (df = 54)	1.504	1.422	0.021	0.021
F statistic (df = 3; 54)	74.450***		66.370***	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 13. OLS, Robust linear, OLS log-linear, and Robust log-linear regression models: 2006 dataset

	Dependent variable			
	Outlet store ratings		Log (Outlet store ratings)	
	OLS	*Robust linear*	*OLS*	*Robust linear*
	(1)	(2)	(3)	(4)
Value	1.649*** (0.495)	0.909** (0.455)		
Selection	1.840*** (0.553)	2.815*** (0.508)		
Help	1.644*** (0.452)	2.015*** (0.279)		
Log (Value)			0.048*** (0.017)	0.045** (0.018)
Log (Selection)			0.092*** (0.021)	0.103*** (0.021)
Log (Help)			0.052*** (0.014)	0.047*** (0.015)
Constant	60.640*** (1.477)	58.960*** (1.357)	4.126*** (0.019)	4.123*** (0.020)
Observations	33	33	33	33
R^2	0.805		0.810	
Adjusted R^2	0.780		0.791	
Residual std. error (df = 29)	1.814	1.292	0.023	0.018
F statistic (df = 3; 29)	38.900***		41.340***	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 14. OLS linear regression models: 2014, 2011, and 2006 datasets

	Dependent variable		
	Outlet store ratings		
	(1)	(2)	(3)
Value	2.975*** (0.503)	3.461*** (0.358)	1.649*** (0.495)
Selection	1.388*** (0.402)	1.760*** (0.371)	1.840*** (0.553)
Help	1.259*** (0.329)	0.987*** (0.264)	1.644*** (0.452)
Constant	60.090*** (1.399)	55.280*** (1.299)	60.640*** (1.477)
Observations	53	58	33
R^2	0.777	0.805	0.801
Adjusted R^2	0.763	0.794	0.780
Residual std. error	1.533 (df = 49)	1.504 (df = 54)	1.814 (df = 29)
F statistic	56.830*** (df = 3; 49)	74.450*** (df = 3; 54)	38.900*** (df = 3; 29)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 15. Log-log linear regression models: 2014, 2011, and 2006 datasets

	Dependent variable		
	Log(Outlet store ratings)		
	(1)	(2)	(3)
Log(Value)	0.110*** (0.019)	0.139*** (0.016)	0.048*** (0.017)
Log(Selection)	0.061*** (0.018)	0.060*** (0.013)	0.092*** (0.021)
Log(Help)	0.037*** (0.012)	0.033*** (0.008)	0.053*** (0.014)
Constant	4.118*** (0.020)	4.052*** (0.019)	4.126*** (0.019)
Observations	53	58	33
R^2	0.759	0.787	0.810
Adjusted R^2	0.745	0.775	0.791
Residual std. error	0.021 (df = 49)	0.021 (df = 54)	0.023 (df = 29)
F statistic	51.550*** (df = 3; 49)	66.370*** (df = 3; 54)	41.340*** (df = 3; 29)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 16. Robust linear regression models (method – 'MM'): 2014, 2011, and 2006 datasets

	Dependent variable		
	Log(Outlet store ratings)		
	(1)	(2)	(3)
Log(Value)	3.042*** (0.517)	3.558*** (0.378)	0.909** (0.455)
Log(Selection)	1.277*** (0.413)	1.791*** (0.391)	2.815*** (0.508)
Log(Help)	1.255*** (0.337)	1.032*** (0.279)	2.015*** (0.415)
Constant	60.280*** (1.436)	54.770*** (1.372)	58.960*** (1.357)
Observations	53	58	33
Residual std. error	1.575 (df = 49)	1.422 (df = 54)	1.393 (df = 29)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 17. Robust log-log linear regression models (method = 'MM'): 2014, 2011, and 2006 datasets

	Dependent variable		
	Log(Outlet store ratings)		
	(1)	(2)	(3)
Log(Value)	0.114*** (0.020)	0.144*** (0.017)	0.045*** (0.018)
Log(Selection)	0.058*** (0.018)	0.058*** (0.013)	0.103*** (0.021)
Log(Help)	0.038*** (0.012)	0.035*** (0.009)	0.047*** (0.015)
Constant	4.118*** (0.021)	4.046*** (0.020)	4.123*** (0.020)
Observations	53	58	33
Residual std. error	0.020 (df = 49)	0.021 (df = 54)	0.018 (df = 29)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.