U.S. DEMAND FOR DAIRY ALTERNATIVE BEVERAGES: HEDONIC METRIC APPROACH

A Dissertation

by

TINGYI YANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Senarath Dharmasena
Committee Members,	Oral Capps, Jr.
	David A. Bessler
	Marco A. Palma
Head of Department,	Mark L. Waller

December 2018

Major Subject: Agricultural Economics

ABSTRACT

The consumption of dairy alternative beverages in the United States has increased rapidly in recent years. Conventional milk seems to be losing its market share to dairy alternative beverages, such as almond and soy milk, mainly because of the taste, health concerns and environmental concerns with these beverages. It is expected that the ongoing competition between conventional milk and dairy alternative beverages will continue to intensify over the next several years.

This work is motivated by the need to take into consideration of intrinsic characteristics and difference of such characteristics when measuring the differentiation of products in demand estimation. Hedonic pricing model is extended by constructing a hedonic matrix for a variety of qualitative attributes of three dairy alternative beverages and four conventional milk categories to model consumer demand for these products in the United States. A two-stage estimation procedure is applied. The first stage is to create a multi-dimensional hedonic space based on the qualitative information available to consumers. The next stage is to allocate the different types of conventional milk and dairy alternative beverages considered in this study into the space and estimate the demand elasticities using attribute distances. Our methodology is advantageous in that it not only solves the price endogeneity problem typically questioned in demand systems estimation but also remarkably reduces the number of parameters estimated. Therefore, it enables a large number of differentiated products to be considered in a demand system.

Nielsen Homescan consumer panel data 2004-2015 is used in this study. The final data set is aggregated to UPC level of products which captures not only enough variation in nutritional variables but also consumer purchase information.

Results show that soy milk has the highest own-price elasticity which is -0.2506. Inelastic demand of all three types of dairy alternative beverages reflects that they are not sensitive to price changes. Soy milk is found to be a substitute for all four types of conventional milk products and vise-versa. Three dairy alternative beverages are complements between each other and four types of conventional milk are substitutes between each other.

DEDICATION

I would like to dedicate this dissertation to my parents, from whom I have gotten unconditional love and support and inherited the great curiosity and passion for learning.

ACKNOWLEDGEMENTS

This dissertation becomes a reality with the kind support and help of many individuals. I would like to extend my sincere thanks to all of them. First and foremost, I would like to show my deepest gratitude to my advisor, Dr. Dharmasena, a very kind, respectable, and responsible professor, for imparting his knowledge and expertise in this study. Without his enlightening instruction and patience, I could not have completed my dissertation.

Secondly, I would like to express my gratitude towards my family for the encouragement, which helped me in completion of this dissertation; My father, who is always by my side and helped me a lot in making this study, and my mother, who served as my inspiration to pursue this undertaking.

Many thanks to my distinguished members of the committee, Dr. Oral Capps, Jr., Dr. David A. Bessler, and Dr. Marco A. Palma for sharing their knowledge and technical know-how.

Last but not least, my thanks and appreciations also go to my classmates and people who have willingly helped me out with their abilities. Thanks to my pre-officemates who are also my best friends Mengqiao Liu and Siwan Song and other close friends in America, Yingqian Yang and Ze Shen, Zhe Li, Jiahui Liu, Xueqin Huang, Guannan Zhao, Suzzanne Achgill, Bob Achgill, Mary Krath, Lela Culpepper, Jacob, Pejman, Rebecca, Chongzhi Zhao, Yang Hu, Ziying Zhao, Yiwen Zhang, Yuhong Lin and her parents, Rui Zhu, Shuoran Du, Jingyun Yang, Xuan Wu, and Xiaowei Zheng. Thanks as well to my friends in China, Ruiyuan Duan, Bei Ma, Yanpu Zhang who always support and encourage me.

CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a dissertation committee consisting of Professor Senarath Dharmasena, Oral Capps, Jr., David A. Bessler and Marco A. Palma of the Department of Agricultural Economics.

The Nielsen Homescan data was collected by the University of Chicago's Kilt's Center for Marketing and was purchased by the Department of Agricultural Economics at Texas A&M University.

All other work conducted for the dissertation was completed by the student independently.

This work was made possible in part by student assistantship from the Department of Agricultural Economics at Texas A&M University. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Agricultural Economics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Background

In the past decade, dairy alternative beverages have gained its market position as a robust competitor for conventional milk in the United States. Consumers have gradually turned away from conventional milk, leading the push towards plant-based milk products. Dairy alternative beverages can be roughly segmented into four divisions: soymilk, almond milk, coconut milk, and others (cashew milk, rice milk, hazelnut milk, hemp milk, and oat milk, etc.).

One of the main characteristics of these products is that even though they are not real milk products with animal origin, they are often fortified with certain nutrients such as protein, vitamin, and calcium to make them comparable with conventional milk. As shown in Copeland and Dharmasena (2016), these beverages are designed not only to quench the thirst but also to provide numerous vitamins, minerals, proteins, and favorable fatty acids.

According to a new report from market research firm Packaged Facts, per capita consumption of conventional milk beverages decreased from year 2000 to 2016 by about 22%. Davis et al. (2010), USDA-ERS (2013) and Copeland and Dharmasena (2016), also showed that per capita consumption of fluid milk in the United States has been dwindling over the past 25 years. However, during the same period, annual sales of dairy alternative products are expected to increase about four times the value of sales in 2021, from $7.37 billion in 2016 to $28 billion in 2021(Figure 1). Even though soy milk and almond milk are the leading categories in U.S. dairy alternative beverage market, they also face heavy competition from other dairy alternative beverages due to the availability of various flavors, low-calorie content, and fortification with multiple nutrients. This availability of an increased variety of dairy alternative beverages dramatically expands consumers' choice spectrum. Dharmasena and Capps (2014) and Copeland and Dharmasena (2016) also argued that although soy milk was the leader in dairy alternative beverage market several years ago, its market share decreased substantially to almond milk, where currently almond milk account for 65% of market share while soy milk has only about 25% of market share. This can be consolidated in figure 2 where the market value worldwide of soy milk is quite stable year by year, but that of almond milk is increasing fast from the year 2013 to 2018 and is expectedly to be valued at approximately $ 5.05 billion in the year 2024.

The shift away from conventional fluid milk products towards dairy alternative beverages in part revolves also around health concerns with a growing number of consumers beginning to believe that plant-based foods are healthier than animal-based foods. Furthermore, because of a growing number of consumers motivated by animal welfare concerns, plant-based beverages have become more preferred than animal-based products. Figure 3 which presents the annual household penetration rate of plant-based beverages in the United States from 2010 to 2016 is showing that the plant-based milk products have penetrated to U.S. household in a continuingly increasing rate from 18% in 2010 to 33% in 2016. Also from figure 4 shows that sales of conventional milk have decreased 13.6 percent ending in 2016, while milk substitutes have increased 6.1 percent.

Even though dairy alternative beverage market has stripped some market share from the conventional fluid milk market, it still possesses only a comparatively small amount of total market share. Conventional milk products are still being consumed in over 90% of the households in U.S. Figure 5 illustrates share of milk sales in the United States in 2016, which shows that skim or low-fat milk are the most popular milk products by consumers with its market share being 55%, and whole milk occupies 26% of the entire milk market. At the same time, almond milk and soy milk account for 7% and 1% market share respectively. Therefore, it is evident that conventional milk products are still the leading force in regular milk and dairy alternative beverage market while almond milk products are gaining popularity among consumers.

Due to the fact that dairy alternative beverages are different from one another in terms of nutritional composition and price, consumers' relative valuations of these beverages with different nutritional content and pricing is of interest to researchers. On May 20, 2016, the United States Federal Drug Administration (US-FDA) announced the new Nutrition Facts label for packaged foods to reflect new scientific information, including the link between diet and chronic diseases such as obesity and heart disease. The new label will make it easier for consumers to make better informed food choices. This action is an indicator that consumers are concerned about the nutritional content in the products they purchase and play an important role in making consumption choices. Also, because of the rising consumption of dairy alternative beverages, research inquiry of consumers' purchasing patterns of these new beverages is of interest as well. From an economic standpoint, consumer expenditure patterns and estimates of own-price and cross-price elasticities as well as expenditure elasticities of those beverages can provide useful information of demand-led growth in dairy alternative beverage market. Plus, this information will be important in order to forecast the future development of the consumption of these beverages. Additionally, this elasticity estimates with regard to dairy alternative beverages will be important for manufactures and advertisers of such beverage products to strategically position dairy alternative beverages in the competitive conventional milk marketplace.

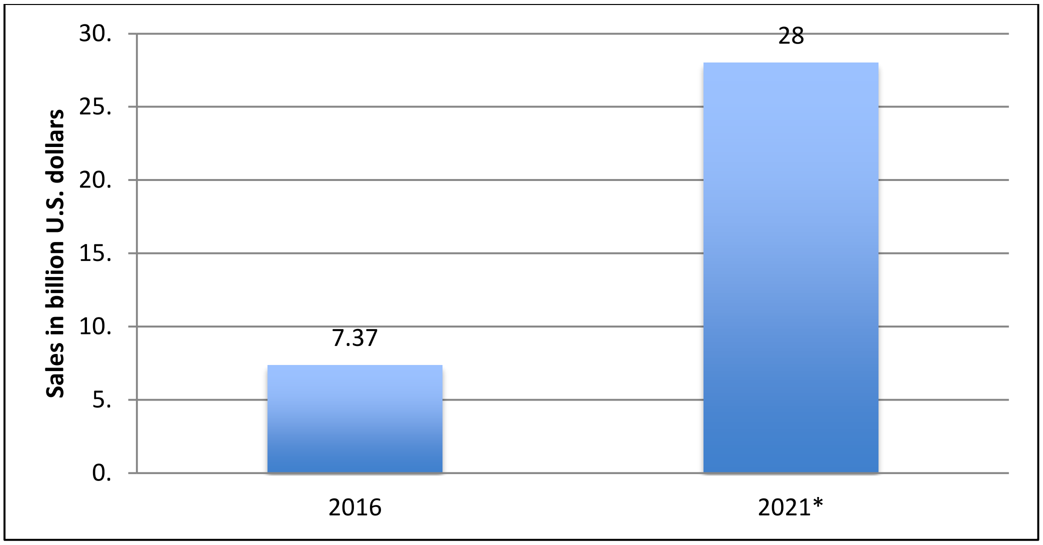


Figure 1 Sales of Dairy Alternative Beverages Worldwide 2016-2022
Adapted from MarketsandMarkets, www.marketsandmarkets.com

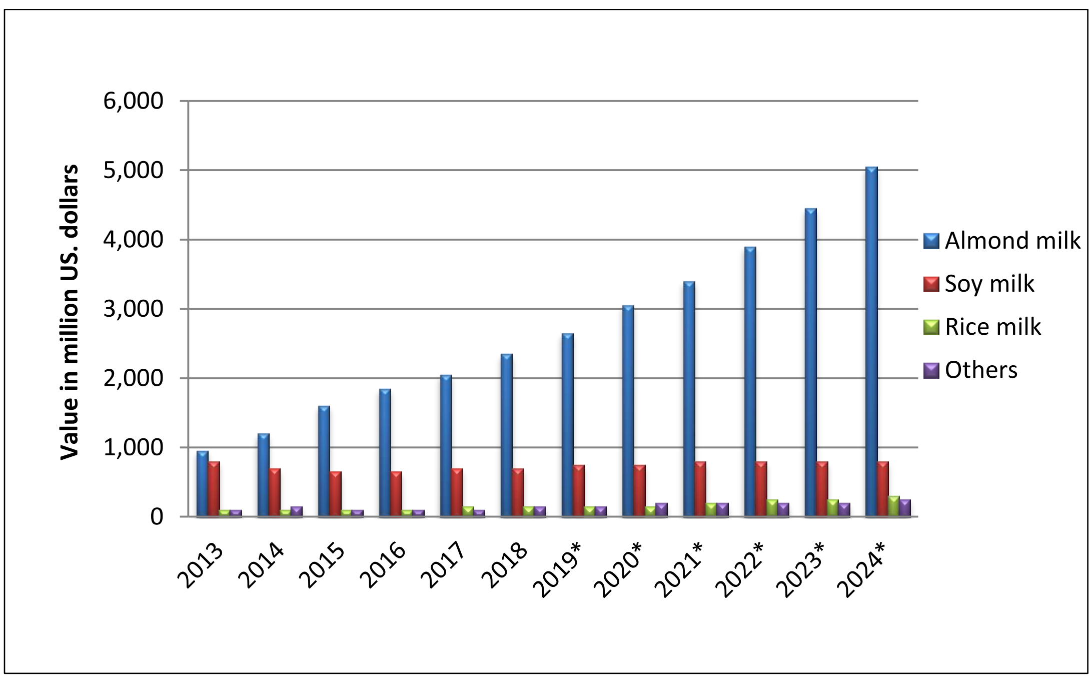


Figure 2 Market Value of Dairy Milk Alternative Beverages Worldwide 2013-2024 by Category
Adapted from Grand View Research; Statista estimates

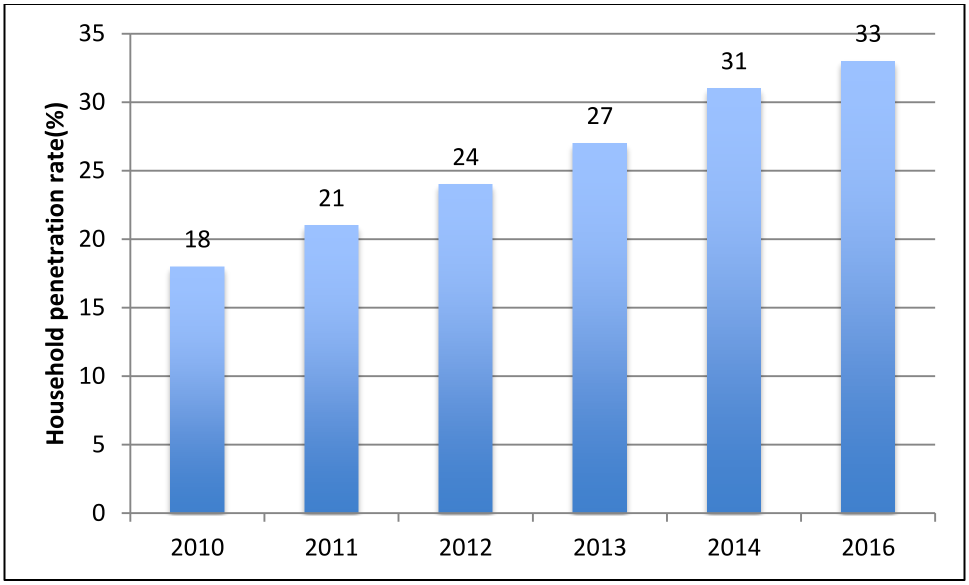


Figure 3 Annual Household Penetration Rate of Plant-based Beverages in the United States 2010-2016

Adapted from Nielsen; WhiteWave Foods; Danone

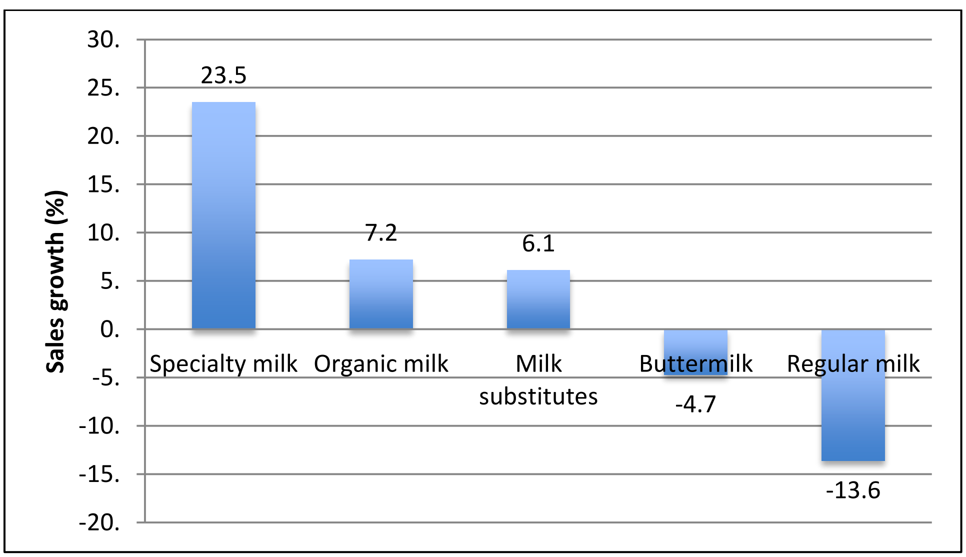


Figure 4 Sales Growth of Milk in the United States in 2016, by Category

Adapted from Bloomberg; Nielsen

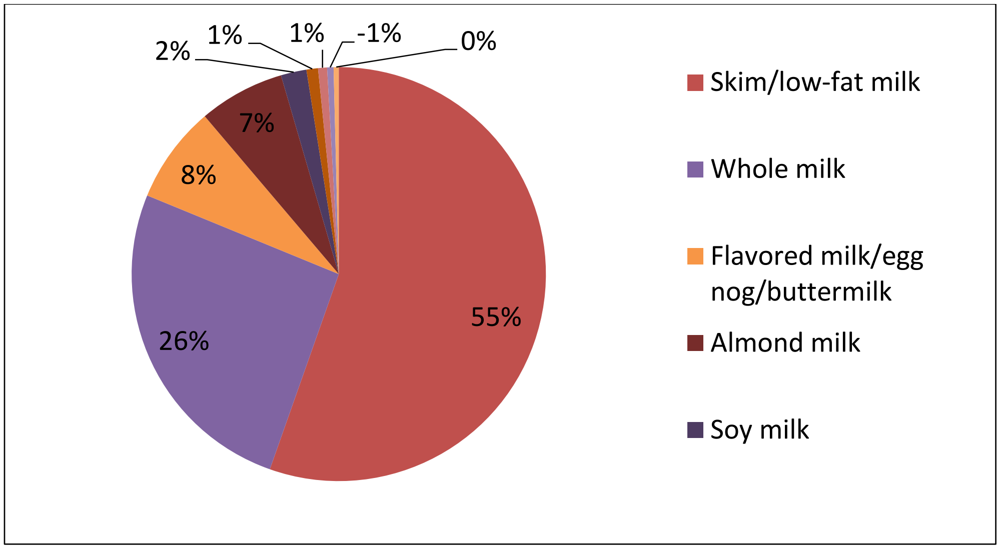


Figure 5 Shares of Milk Sales in the United States in 2016, by Category
Adapted from Information Resources, Incorporated (IRI); Dairy Foods, Progressive Grocer Magazine, July 2016, page 64

Literature Review

Many studies exist in the extant literature focusing on consumer demand in estimating price and expenditure elasticities of food products. In conventional demand models, such as the almost ideal demand system (AIDS) (Deaton and Muellbauer, 1980) and the Rotterdam model (Theil, 1965), the relationship between prices and market shares is exploited to estimate the own-price, cross-price, and expenditure elasticities (or income elasticities). Several studies in the extant literature apply those traditional models to study demand and market competitiveness of dairy alternative beverages in the United States. Dharmasena and Capps (2014) estimated the demand for soymilk, white milk, and flavored milk. Copeland and Dharmasena (2016) analyzed demand for dairy alternative beverages and the effect of increased demand for those products on dairy farmers' welfare. Li and Dharmasena (2016) also estimated demand for coconut milk.

However, traditional demand models are based on the assumption that consumers' utility is gain through the quantity of a specific good they consume, without considering the intrinsic properties of a particular good which distinguishes it from other similar goods. The random utility model, which is based on the characteristics approach to estimating demand, has been widely used as an alternative model to conventional demand model. The Random Utility Model is extended from simulated maximum likelihood approach of Berry et al. (1995). It has three basic assumptions representing a distinct break with traditional demand models, and two of them are related to this work. It assumes that the product attributes (or characteristics) with which the good possess give rise to utility (Lancaster, 1966) and not just the quantity of the consumed good. Hence, the total amount of utility a consumer enjoys from his purchases of products depends on the total amount of product characteristics that consumer intrinsically looked at in purchasing the product.

According to Triplett's (1986), characteristics are 'homogeneous economic variables' that are 'packaged' or 'bundled' into a specific product. Another assumption is that, generally a good will have more than one characteristic, and many characteristics will be shared by more than one good (Lancaster, 1966). Many studies contribute to the theoretical development and empirical application of the random utility model. Lancaster (1966) who thoroughly discussed the argued that the characteristics of goods are part of the consumer's utility function and preferences depend on the measure of each desired characteristic. Rosen (1974) also showed that consumers maximize utility by selecting products that maximize the sum of utilities derived from each attribute. Ladd and Suvannunt (1976) developed their theoretical model called consumer goods characteristics model (CGCM) and compared it with Lancaster's analysis and tested their model using data from the automobile industry. From that research, we can conclude that the total amount of utility a consumer enjoys from his purchases of products depends on the total amounts of product characteristics purchased rather than only the quality of the products.

However, even though RUMs reflect characteristic differences in elasticities, they are very complicated computationally and difficult to estimate (Hendel 1999; Nevo 2001; Chan 2006; Gulseven and Wohlgenant, 2010). Given this issue, Pinske, Slade, and Brett (2002) developed a distance metric (DM) approach which uses spatial distance to the desired characteristic to estimate price elasticities for different products. Compared to the RUM approach which requires a simulation process, the DM approach is more straightforward to apply. Moreover, it is flexible enough to characterize the substitution patterns between differentiated products (Gulseven and Wohlgenant, 2010). Rojas and Peterson (2008) apply this method to retail beer market. They selected alcohol content as the primary quality measure of beer products supplemented by other different distance combinations. However, one of the prominent weaknesses of their work is that the distance measurement they construct is ambiguous and it is built on prior judgments about the data, specifically in selecting the base category of the product in developing the distance matrix.

In order to remedy these limitations, specifically the selection on arbitrary base category in the construction of the distance matrix, Gulseven and Wohlgenant (2010) proposed a new approach, hedonic metric approach, to examine and estimate the demand for retail milk, showing that this approach alleviates the embedded ambiguity in selecting the base category problem in the DM model. Another notable advantage of this model is that it significantly reduces the number of parameters to be estimated, thereby making it possible to estimate large number of differentiated products in a single demand system.

The first step of hedonic matrix approach is to estimate the hedonic pricing models. There are numerous studies available in the extant literature that applied this method to analyze the effects of attributes or characteristics on prices of differentiated products. As shown by Epple (1987), in the empirical investigation of hedonic models, one issue of interest is to determine how the price of a unit of the commodity varies with the set of characteristics or attributes it possesses. The other interesting subject is to estimate the (underlying) demand and supply functions for the characteristics, such as the work by Palmquist (1984) who estimated the demand for housing based on the characteristic approach. The basic concept of hedonic pricing models is that products are wanted because of the utilities they provide. The utilities provided depend upon the product characteristics. Hence, the total amount of utility a consumer enjoys from his purchases of products depends upon the total amounts of product characteristics purchased.

Similar to RUMs, hedonic pricing models are also a commonly used method to recover consumer preferences in a differentiated product market. They are based on different assumptions where stochastic preference shocks are considered in the Random Utility Models, but omitted in hedonic pricing models. Rosen (1974) defined hedonic prices as ‘the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specified amounts of characteristics associated with them’. This method is applied after the work of Griliches (1961) who analyzed quality-adjusted measures of automobile prices. The application of this model has its origins in agricultural economics. For example, Waugh (1928, 1929) estimated relations between the wholesale prices of fresh vegetables and their characteristics. This method is still widely used in recent agricultural economics studies. For instance, Bryant (2012) developed a hedonic pricing model for post-extracted algae residue used for assessing the economic feasibility of an algal production enterprise. Even though hedonic pricing method is applied widely in the area of agricultural commodities and other differentiated products, little work has been done to examine the link between the quality attributes and price differentials to estimate consumer demand for dairy alternative beverages. There have been studies based on Lancaster's (1966) hedonic framework, such as models applied in the research of Gould and Lin (1994) and Huang (1996). However, the primary motivation of these models is to estimate demand for qualitative characteristics based on product elasticities, whereas our work is aiming to estimate the demand for the products based on their qualitative characteristics. Gulseven and Wohlgenant (2015) used the method to estimate own-price, cross price and expenditure elasticities of four milk types, whole, reduced fat, low fat and skim and soymilk based on quality attributes and hedonic prices. But the milk product types included in their study are limited and mingled with dairy alternative beverages since soymilk is not commonly categorized into conventional milk categories.

Research Questions and Objectives

Given the lack of research on dairy alternative beverage market and application of hedonic metric approach to analyze demand of milk alternative beverages, we plan to do the following: (1) develop and estimate linear and semi-log hedonic pricing models for almond milk, soy milk, rice milk and four types of conventional milk (1% fat, 2% fat, fat-free milk, and whole milk); (2) construct attribute-space hedonic matrix for each beverage category at brand-level for a pre-identified set of brands; (3) estimate consumer demand for the aforementioned beverage categories using attribute space hedonic model augmented Barten Synthetic model (Barten, 1993); and (4) explore the price sensitivities, substitute and complement effects among three dairy alternative beverages and four types of milk products, thus providing information for manufacturers, retailers, advertising companies, nutritionists and other stakeholders for strategic decision making.

The organization of this dissertation is as follows. The next chapter focuses on introducing the methodology applied in this work. The major part of methodology is separated into two stages. In the first stage, we estimate the hedonic pricing models, where prices are defined as a function of products' qualitative characteristic. In the second stage we estimate expenditure elasticities and own-price and cross-price elasticities using the hedonic matrix augmented Barten synthetic model. In the third chapter, we discuss the data, how the data are acquired and organized for this work. The fourth chapter shows the estimated results from hedonic pricing models and Barten synthetic demand model. Chapter five offers concluding remarks. In the sixth chapter, study limitations and some interesting future research topics are discussed.

CHAPTER II

METHODOLOGY

Some parts in this chapter follow the hedonic metric method proposed in Gulseven and Wohlgenant (2015). It is based on a two-stage estimation technique. First we discuss the hedonic pricing models from which we obtain the marginal value (or shadow price) of each quality attribute available in the product. Then, a hedonic matrix is created based on the hedonic distances of different quality attributes. These measures are based on pair-wise comparisons of Euclidean distances, where the amount of each characteristic in the product is weighted by its shadow price. After each differentiated product considered in this study is allocated into the hedonic space, these distances are used to reparameterize the price coefficients in estimating demand. As such, hedonic matrix augmented Barten synthetic demand system is estimated in this study.

Hedonic Pricing Models

Hedonic pricing models assume that the consumer maximizes utility by selecting products that maximize the sum of the utilities derived from each attribute (Rosen, 1974). Therefore, the price of each beverage in this study can be explained by the set of attributes of the product. We can define this set as $x = [x_i, \ldots, x_j]$, the functional relationship between the price of a good and its characteristics vector x can be generally shown as $p = f(x) + \mu$, where μ is the error vector, and p is the observed price. If the relationship between prices and attributes is assumed to be linear, price of a good *i* can be derived as the sum of the attribute values (Ladd and Suvannunt, 1976). Thus the total value of each attribute is equal to the quantity of the attribute multiplied by the implicit price of that attribute (Gulseven and Wohlgenant, 2015).

(1) $P_i = \beta_0 + \sum_j \beta_j A_{ij} + \sum_k D_k X_{ik} + \varepsilon_i, \;\; i = 1,2, \dots ,7$

(2) $\ln(P_i) = \beta_0 + \sum_j \beta_j A_{ij} + \sum_k D_k X_{ik} + \varepsilon_i, \;\; i = 1,2, \dots ,7$

where A_{ij} is the amount of nutritional attribute j contained in product i. X_{ik} is other factors that might affect prices.

The implicit prices or shadow prices of linear hedonic regression can be shown as:

(3) $\frac{\partial P_i}{\partial A_{ij}} = \frac{\partial f(x)}{\partial A_{ij}} = \beta_j \;\; \forall i, j$

(4) $\frac{\partial P_i}{\partial X_{ik}} = \frac{\partial f(x)}{\partial D_{ik}} = D_k \;\; \forall i, k$

Marginal effect of semi-log hedonic pricing model is derived as follows. First, solve for P_i from equitation (2):

(5) $P_i = e^{(\beta_0 + \sum_j \beta_j A_{ij} + \sum_k D_k X_{ik} + \varepsilon_i)}$

then differentiate equation (5) to get the marginal effect of A_{ij} and X_{ik}

(6) $\frac{\partial P_i}{\partial A_{ij}} = \beta_j e^{\left(\beta_0 + \sum_j \beta_j A_{ij} + \sum_k D_k X_{ik} + \varepsilon_i\right)} = \beta_j P_i \; \forall i, j$

(7) $\frac{\partial P_i}{\partial X_{ik}} = D_k e^{\left(\beta_0 + \sum_j \beta_j A_{ij} + \sum_k D_k X_{ik} + \varepsilon_i\right)} = D_k P_i \; \forall i, k$

The two equations (1) and (2) outline the relationship between different categories of explanatory variables and the dependent variable P_i, the price for the ith product. Specifically, as shown in equation (1), prices of the products are intrinsically represented by hedonic attributes of the products in this study. P_i is the monthly average price recorded in Nielsen database by different Universal Product Codes (UPC) that have been purchased from the year 2004 to 2015. Those nutritional attributes include calories, fat, fiber, protein, calcium, Vitamin A, etc. If the price attribute relationship is assumed to be in a semi-log form (Nimon and Beghin 1999), then instead of price, the log-price of the product is defined regarding attributes as is shown in equation (2). Similarly, as shown in equation (1), P_i is the monthly average prices of a beverage from year 2004 to 2015. The implicit prices are the coefficients to be estimated which are represented by β_j and D_k. Different from previous research, we also take into consideration of time effects on the price of each year by adding yearly dummies into the model. Therefore, all attributes are separated into nutritional attributes and other related attributes that might affect prices of the products including package size, values of the multi-package, brands, coupon, and yearly dummies, etc. β_0 denotes the intercept and ε_i represents the stochastic error term. If we specify all the variables in the model, then the linear and log hedonic pricing model in this research can be demonstrated as equation (5) and (6) respectively:

(8) $P_i = \beta_{\mathrm{i}0} + \beta_{i1}Calories + \beta_{i2}fat + \beta_{i3}VitaminA + \beta_{i4}calcium + \beta_{i5}VitaminD + \beta_{i6}fiber + \beta_{i7}protien + \beta_{i8}brands + +\beta_{i9}coupon + \beta_{i10}deal + \sum_{k=1}^{n}\beta_{i1k}package_{size} + \sum_{m=1}^{n}\beta_{i2m}multi + \sum_{t=1}^{14}\beta_{i3t}year + \varepsilon_i$, $i = 1,2,\ldots,7$

(9) $lnP_i = \beta_{i0} + \beta_{i1}Calories + \beta_{i2}fat + \beta_{i3}VitaminA + \beta_{i4}calcium + \beta_{i5}VitaminD + \beta_{i6}fiber + \beta_{i7}protien + \beta_{i8}brands + +\beta_{i9}coupon + \beta_{i10}deal + \sum_{k=1}^{n}\beta_{i1k}package_{size} + \sum_{m=1}^{n}\beta_{i2m}multi + \sum_{t=1}^{14}\beta_{i3t}year + \varepsilon_i$, $i = 1,2,\ldots,7$

where $\varepsilon_i \sim N(0, \Sigma^*)$, Σ^* is an *n x n* singular covariance matrix. Except for nutritional variables, the hedonic variables include coupon, deal, package size dummies, multi-pack dummies and yearly dummies, but their values vary for different milk types. *i* represents soy milk, almond milk, rice milk, 2% low-fat milk, 1% low-fat milk, whole milk and fat-free milk; *k* is the value of package sizes; *m* is the units purchased together; *t* is regarding to the time series from year 2004 to 2015; *n* means that for different product, their values and numbers of package size and multi-pack dummies are different as shown from table 1-7.

Table 1 Description of Dummy Variables of Package Size and Multi-Pack: Almond Milk

Dummy variables	Labels	Description
$D_{\text{pkge\_size1}}$	package_size1	8 oz.
$D_{\text{pkge\_size2}}$	package_size2	10 oz.
$D_{\text{pkge\_size3}}$	package_size3	10 oz. <size1_amount<11 oz.[a]
$D_{\text{pkge\_size4}}$	package_size4	11 oz.
$D_{\text{pkge\_size5}}$	package_size5	12 oz.
$D_{\text{pkge\_size6}}$	package_size6	16 oz.
$D_{\text{pkge\_size7}}$	package_size7	32 oz.
$D_{\text{pkge\_size8}}$	package_size8	48 oz.
$D_{\text{pkge\_size9}}$	package_size9	64 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	2 packs
D_{multi4}	multi_4	4 packs
D_{multi5}	multi_6	6 packs
D_{multi7}	multi_7	12 packs
D_{multi8}	multi_8	18 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. “multi”>1) is total units for a product;
- *size1_units* is the unit of measure. For example, “size1_amount” might be “16.0”,

and” size1_units” might be “OZ.” ;

Because all brands of almond milk products included in this study are national brands, we decided not to include brand dummy in the hedonic regressions of almond milk.

[a] For package_size dummy variables that are in ranges, for example 10 oz.<size1_amount<11oz., they are created because there are many package sizes that are not integers and for different types of beverages, the values vary a lot. In order to make the package sizes comparable from one beverage to another, we created some package size dummies that are in ranges.

Table 2 Description of Dummy Variables of Package Size and Multi-Pack: Rice Milk

Dummy variables	Labels	Description
D_{pkge_size1}	package_size1	11 oz.
D_{pkge_size2}	package_size2	12 oz.
D_{pkge_size3}	package_size3	14 oz.
D_{pkge_size4}	package_size4	16 oz.
D_{pkge_size5}	package_size5	32 oz.
D_{pkge_size6}	package_size6	48 oz.
D_{pkge_size7}	package_size7	64 oz.
D_{pkge_size8}	package_size8	128 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	12 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. “multi”>1) is total units for a product;
- *size1_units* is the unit of measure. For example, “size1_amount” might be “16.0”, and” size1_units” might be “OZ.” ;

Table 3 Description of Dummy Variables of Package Size and Multi-Pack: Soy Milk

Dummy variables	Labels	Description
$D_{\text{pkge\_size1}}$	package_size1	8 oz.
$D_{\text{pkge\_size2}}$	package_size2	8 oz. <size1_amount<16 oz.[a]
$D_{\text{pkge\_size3}}$	package_size3	10 oz.
$D_{\text{pkge\_size4}}$	package_size4	10 oz. <size1_amount<11 oz.[a]
$D_{\text{pkge\_size5}}$	package_size5	11 oz.
$D_{\text{pkge\_size6}}$	package_size6	12 oz.
$D_{\text{pkge\_size7}}$	package_size7	15 oz.
$D_{\text{pkge\_size8}}$	package_size8	15 oz. <size1_amount<16 oz.[a]
$D_{\text{pkge\_size9}}$	package_size9	16 oz.
$D_{\text{pkge\_size10}}$	package_size10	32 oz.
$D_{\text{pkge\_size11}}$	package_size11	32 oz. <size1_amount<48oz[a]
$D_{\text{pkge\_size12}}$	package_size12	48 oz.
$D_{\text{pkge\_size13}}$	package_size13	64 oz.
$D_{\text{pkge\_size14}}$	package_size14	128 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	2 packs
D_{multi3}	multi_3	3 packs
D_{multi4}	multi_4	4 packs
D_{multi5}	multi_5	6 packs
D_{multi6}	multi_6	12 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. "multi">1) is total units for a product;
- *size1_units* is the unit of measure. For example, "size1_amount" might be "16.0",

and" size1_units" might be "OZ." ;

[a] For package_size dummy variables that are in ranges, for example 8oz.<size1_amount<16oz., they are created because there are many package sizes that are not integers and for different types of beverages, the values vary a lot. In order to make the package sizes comparable from one beverage to another, we created some package size dummies that are in ranges.

Table 4 Description of Dummy Variables of Package Size and Multi-Pack: 1% Milk

Dummy variables	Labels	Description
$D_{\text{pkge\_size1}}$	package_size1	8 oz.
$D_{\text{pkge\_size2}}$	package_size2	10 oz.
$D_{\text{pkge\_size3}}$	package_size3	10 oz. <size1_amount<11 oz.[a]
$D_{\text{pkge\_size4}}$	package_size4	12 oz.
$D_{\text{pkge\_size5}}$	package_size5	14 oz.
$D_{\text{pkge\_size6}}$	package_size6	16 oz.
$D_{\text{pkge\_size7}}$	package_size7	32 oz.
$D_{\text{pkge\_size8}}$	package_size8	52 oz.
$D_{\text{pkge\_size9}}$	package_size9	52<size1_amount<64 oz.[a]
$D_{\text{pkge\_size10}}$	package_size10	64 oz.
$D_{\text{pkge\_size11}}$	package_size11	94 oz.
$D_{\text{pkge\_size12}}$	package_size12	96 oz.
$D_{\text{pkge\_size13}}$	package_size13	97 oz.
$D_{\text{pkge\_size14}}$	package_size14	128 oz.
$D_{\text{pkge\_size15}}$	package_size15	192 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	2 packs
D_{multi3}	multi_3	3 packs
D_{multi6}	multi_6	6 packs
D_{multi7}	multi_7	12 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. “multi”>1) is total units for a product;
- *size1_units* is the unit of measure. For example, “size1_amount” might be “16.0”,

and” size1_units” might be “OZ.” ;

[a] For package_size dummy variables that are in ranges, for example 10oz.<size1_amount<11oz., they are created because there are many package sizes that are not integers and for different types of beverages, the values vary a lot. In order to make the package sizes comparable from one beverage to another, we created some package size dummies that are in ranges.

Table 5 Description of Dummy Variables of Package Size and Multi-Pack: Fat-free Milk

Dummy variables	Labels	Description
$D_{\text{pkge\_size1}}$	package_size1	size1_amount<8 oz.[a]
$D_{\text{pkge\_size2}}$	package_size2	8 oz.
$D_{\text{pkge\_size3}}$	package_size3	10 oz.
$D_{\text{pkge\_size4}}$	package_size4	12 oz.
$D_{\text{pkge\_size5}}$	package_size5	14 oz.
$D_{\text{pkge\_size6}}$	package_size6	16 oz.
$D_{\text{pkge\_size7}}$	package_size7	20 oz.
$D_{\text{pkge\_size9}}$	package_size9	32 oz.
$D_{\text{pkge\_size10}}$	package_size10	32<size1_amount<52 oz.[a]
$D_{\text{pkge\_size11}}$	package_size11	52 oz.
$D_{\text{pkge\_size12}}$	package_size12	52<size1_amount<64 oz.[a]
$D_{\text{pkge\_size13}}$	package_size13	64 oz.
$D_{\text{pkge\_size14}}$	package_size14	94 oz.
$D_{\text{pkge\_size15}}$	package_size15	96 oz.
$D_{\text{pkge\_size16}}$	package_size16	128 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	2 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. "multi">1) is total units for a product;
- *size1_units* is the unit of measure. For example, "size1_amount" might be "16.0",

and" size1_units" might be "OZ." ;

[a] For package_size dummy variables that are in ranges, for example size1_amount<8 oz., they are created because there are many package sizes that are not integers and for different types of beverages, the values vary a lot. In order to make the package sizes comparable from one beverage to another, we created some package size dummies that are in ranges.

Table 6 Description of Dummy Variables of Package Size and Multi-Pack: Whole Milk

Dummy variables	Labels	Description
$D_{\text{pkge\_size1}}$	package_size1	package_size< 8 oz.[a]
$D_{\text{pkge\_size2}}$	package_size2	8 oz.
$D_{\text{pkge\_size3}}$	package_size3	10 oz.
$D_{\text{pkge\_size4}}$	package_size4	10 oz. <size1_amount<11 oz.[a]
$D_{\text{pkge\_size5}}$	package_size5	12 oz.
$D_{\text{pkge\_size6}}$	package_size6	14 oz.
$D_{\text{pkge\_size7}}$	package_size7	16 oz.
$D_{\text{pkge\_size8}}$	package_size8	20 oz.
$D_{\text{pkge\_size9}}$	package_size9	24 oz.
$D_{\text{pkge\_size10}}$	package_size10	32 oz.
$D_{\text{pkge\_size11}}$	package_size11	32 oz. <size1_amount<52 oz.[a]
$D_{\text{pkge\_size12}}$	package_size12	52 oz.
$D_{\text{pkge\_size13}}$	package_size13	52 oz. <size1_amount<64 oz.[a]
$D_{\text{pkge\_size14}}$	package_size14	64 oz.
$D_{\text{pkge\_size15}}$	package_size15	96 oz.
$D_{\text{pkge\_size16}}$	package_size16	128 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	2 packs
D_{multi3}	multi_3	3 packs
D_{multi5}	multi_5	5 packs
D_{multi6}	multi_6	6 packs
D_{multi7}	multi_7	9 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. “multi”>1) is total units for a product;
- *size1_amount* is package size (numeric size of the product)

a For package_size dummy variables that are in ranges, for example size1_amount<8 oz., they are created because there are many package sizes that are not integers and for different types of beverages, the values vary a lot. In order to make the package sizes comparable from one beverage to another, we created some package size dummies that are in ranges.

Table 7 Description of Dummy Variables of Package Size and Multi-Pack: 2% Milk

Dummy variables	Labels	Description
$D_{\text{pkge\_size1}}$	package_size1	package_size <8 oz.[a]
$D_{\text{pkge\_size2}}$	package_size2	8 oz.
$D_{\text{pkge\_size3}}$	package_size3	8 oz. <size1_amount<10 oz.[a]
$D_{\text{pkge\_size4}}$	package_size4	10 oz.
$D_{\text{pkge\_size5}}$	package_size5	10 oz. <size1_amount<11 oz.[a]
$D_{\text{pkge\_size6}}$	package_size6	11 oz.
$D_{\text{pkge\_size7}}$	package_size7	12 oz.
$D_{\text{pkge\_size8}}$	package_size8	14 oz.
$D_{\text{pkge\_size9}}$	package_size9	16 oz.
$D_{\text{pkge\_size10}}$	package_size10	20 oz.
$D_{\text{pkge\_size11}}$	package_size11	24 oz.
$D_{\text{pkge\_size12}}$	package_size12	32 oz.
$D_{\text{pkge\_size13}}$	package_size13	32 oz. <size1_amount<52 oz.[a]
$D_{\text{pkge\_size14}}$	package_size14	52 oz.
$D_{\text{pkge\_size15}}$	package_size15	64 oz.
$D_{\text{pkge\_size16}}$	package_size16	94 oz.
$D_{\text{pkge\_size17}}$	package_size17	96 oz.
$D_{\text{pkge\_size18}}$	package_size18	97 oz.
$D_{\text{pkge\_size19}}$	package_size19	128 oz.
D_{multi1}	multi_1	1 (single serve)
D_{multi2}	multi_2	2 packs
D_{multi3}	multi_3	3 packs
D_{multi6}	multi_6	6 packs
D_{multi7}	multi_7	12 packs

- *multi* its value represents number of units in multipack;
- *multi packs* (i.e. “multi”>1) is total units for a product;
- *size1_amount* is package size (numeric size of the product)

[a] For package_size dummy variables that are in ranges, for example size1_amount<8 oz., they are created because there are many package sizes that are not integers and for different types of beverages, the values vary a lot. In order to make the package sizes comparable from one beverage to another, we created some package size dummies that are in ranges.

The reason why we include the three types of dairy milk alternative beverages and four types of milk products is as follows. First, soy milk and almond milk are two categories that have the largest market shares in the milk alternative beverage market. Rice milk and coconut milk are another two categories that are becoming popular and in recent times. Figure 6 illustrates the sales value of rice milk in the United States from 2013 to 2016 and provides a forecast up to 2024. In 2016, it is evident that the sales of rice milk show an increasing trend and are expected to grow to about $275 million by 2024. Due to lack of data with regard to coconut milk purchases, this category cannot be included in this study. It is well-known that whole milk, 2% low-fat milk, 1% low-fat milk and fat-free milk (or skim milk) is the four most common types in the conventional U.S. dairy milk market, hence included in this study.

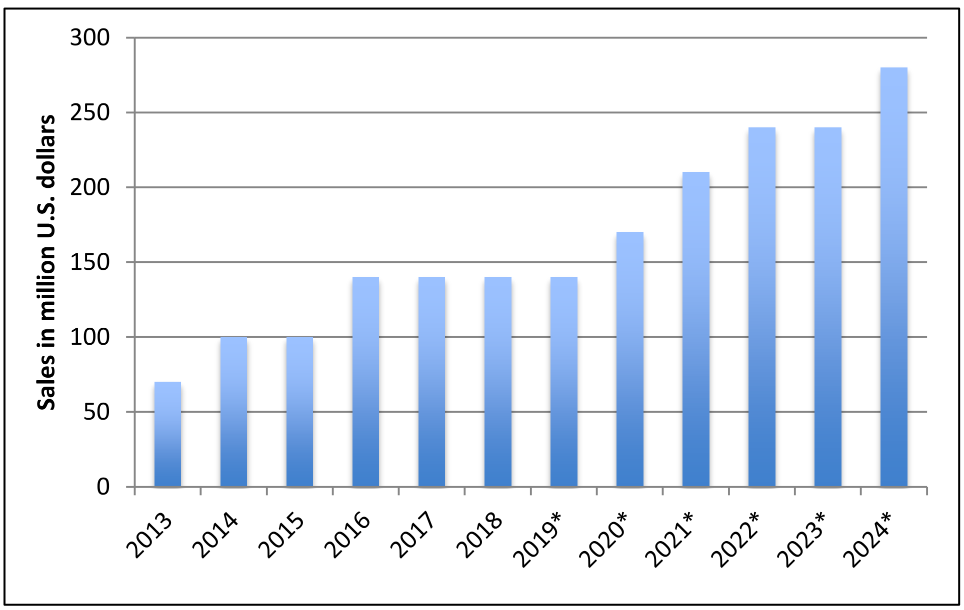


Figure 6 Sales Value of Rice Milk in the United States 2013-2016 and Forecasts to 2024
Adapted from Grand View Research; Statista; Statista estimate

Besides estimating the impact of attributes of products on prices, it is also meaningful to calculate the value added for each attribute by multiplying the implicit price by the attribute quantity which is demonstrated in equation (6) and (7). For each beverage i, the value added from attribute j is $v_{ij} = x_{ij}\beta_j$. If use the log form, then the implicit price of the attribute is calculated by multiplying the coefficients on attributes with the price of the products. The value added term also accounts for the price of the product $v_{ij} = x_{ij}\beta_j P_i$. The semi-log form implies that the same amount of qualitative factor can have a higher value if it is located in a product with a higher retail price. The continuous hedonic distance matrix is calculated based on the value added terms from hedonic regressions through combining the sum of price-weighted attribute distances and rescaling them to be between 0 and 1 to make those distances comparable. Also, it is defined that two products are nearest neighbors if they are located next to each other in the hedonic space. The hedonic metric space thus would show as the form in figure 7. The own-price and cross-price effects in the Barten synthetic model now are characterized as equation (10) and (11):

(10) $\gamma_{ij} = c_h d_{ij}^h + c_{nn} d_{ij}^{nn}$,

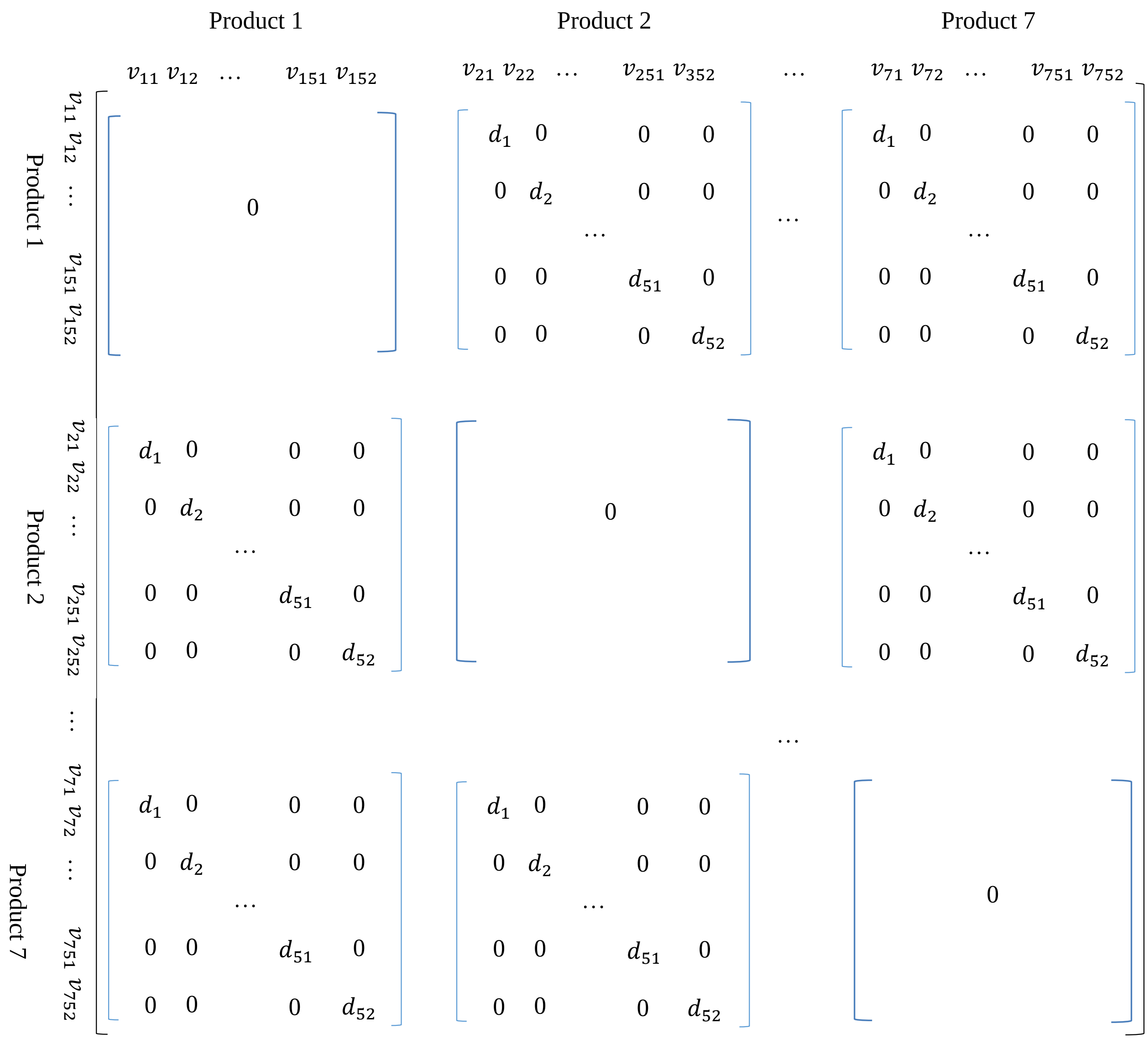
where d_{ij}^h is the hedonic distance, d_{ij}^{nn} is the nearest neighbor dummy matrix.

(11) $\gamma_{ii} = \alpha_0 + \alpha_1 x_i^s + \alpha_2 x_i^c$,

where x_i^s refers to the market share and x_i^c is the closeness index of product i.

Putting these into the original Barten synthetic model will give rise to the Hedonic Metric augmented Barten synthetic model.

From figure 7, we can get a general view of how hedonic attribute space is constructed. Every product has 52 value added terms even though some are equal to zero due to their different values and numbers of the dummy variables other than nutritional variables. The diagonal is zero in the matrix because the distance between the same value added terms is zero. Accordingly, in each block (for every product) in the matrix, it has 52 different attribute distances with another product. For simplicity, we denote those distances as d1 to d52 in each block but they have different values. Later we will discuss in detail how both linear and log hedonic distance matrixes are derived.



v_{ij} is value added from attribute j in product i, $i = 1,2, \dots ,7$ $j = 1,2, \dots ,52$
d_j is the distances of attribute j between two products

Figure 7 Basic Form of Hedonic Attribute Space

Hedonic Distance Measurement

The consideration of putting products in a space and measuring their difference based on the locations in the space can be traced back to a famous model, namely, Hotelling Model, which is developed by Hotelling (1929) who attempted to explain the Bertrand Paradox[1] in oligopolistic competition. He suggested that the difference of two products can be considered as the distance between two companies that produce the products. The closer the two companies, the similar the products they produce. This is the first argument taking into consideration of distance to measure the difference between two products. Therefore, putting products in to a space and measure their difference based on distances which is the main idea in hedonic distance measurement has its origins and can be justified. In this study, hedonic distances are calculated based on the concept of squared difference. There are many distance functions one could use to calculate distances in a data space, in which the most commonly used is the Euclidean distance. The "distance" referred to is the distance between observed samples in the N-dimensional attribute space. Hedonic distances are obtained as the square root of the sum of the squared differences of valued added terms of all attributes. According to Dohnal (2004), a matrix space is defined as a pair $\mathcal{M} = (\mathcal{D}, d)$. The set D denotes the domain (universe) of valid objects (elements, points) of the metric space. A finite subset of the domain $\mathcal{D}$, $X \subseteq \mathcal{D}$, is usually what we call the database. The function

$$d: \mathcal{D} \times \mathcal{D} \longmapsto \mathbb{R}$$

denotes a measure of distance between objects. The smaller the distance is, the closer or more

[1] Suppose there are two firms and they produce identical goods. The Nash Equilibrium is that firms produce at $P_1 = P_2 = MC$, that is the product is selling at zero economic profit. Bertrand's result is paradoxical because if the number of firms goes from one to two, the price decrease from the monopoly price to the competitive price and stays at the same level as the number of firms increase further. This is unrealistic because in reality, markets featuring a small number of firms with market power typically charge a price in excess of marginal cost.

similar the objects are. The distance functions are metric and satisfy following properties:

(12) $\forall x, y \in \mathcal{D}, d(x, y) \geq 0$ Non-negativity

(13) $\forall x, y \in \mathcal{D}, d(x, y) = d(y, x)$ Symmetry

(14) $\forall x \in \mathcal{D}, d(x, x) = 0$ Reflexivity

(15) $\forall x, y \in \mathcal{D}, x \neq y \Rightarrow d(x, y) > 0$ Positiveness

The hedonic distance matrix satisfies all the above properties. Consider a collection of *n* points in a *d*-dimensional Euclidean space, ascribed to the columns of matrix $\boldsymbol{X} \in \mathbb{R}^{d \times n}$, $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2} \cdots, \boldsymbol{x_n}]$, $\boldsymbol{x_i} \in \mathbb{R}^d$, Then the squared distance between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ is given as

(16) $d_{ij} = \left[(\boldsymbol{x_i} - \boldsymbol{x_j})^T (\boldsymbol{x_i} - \boldsymbol{x_j})\right]^{1/2}$

After computing for all i, j and arranging those in a matrix in an order identical to that obtained from matrix cross products, this process results in a symmetric dissimilarity matrix of Euclidean distances, *D*. For *n* cases, *D* will be $n \times n$. Because we have seven products considered in this study, the dimension of hedonic distances is 7×7. This matrix has zeros along the main or principal diagonal (because the distance between a vector and itself is zero) and has units identical to the input data. Table 8 and 9 shows the hedonic distance matrixes calculated based on value added terms acquired from linear hedonic pricing models and log hedonic pricing models, respectively. Table 10 and 11 represents the continuous hedonic distance matrixes which are generated by combining the sum of these price-weighted attribute distances and rescaling them to be between 0 and 1. The formula is as follows:

(17) $d_{ij} = \frac{1}{1 + \sqrt{\left[(\boldsymbol{x_i} - \boldsymbol{x_j})^T (\boldsymbol{x_i} - \boldsymbol{x_j})\right]}}$

Table 8 Linear Hedonic Distance Matrix

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	0	0.2612	0.5403	0.7353	0.5195	0.3875	1.1250
soy milk	0.2612	0	0.3793	0.5914	0.3508	0.2014	1.0922
rice milk	0.5403	0.3793	0	0.7774	0.5729	0.3584	0.9929
2% milk	0.7353	0.5914	0.7774	0	0.5839	0.6774	1.3967
1% milk	0.5195	0.3508	0.5729	0.5839	0	0.4125	1.0276
fat-free milk	0.3875	0.2014	0.3584	0.6774	0.4125	0	1.1638
whole milk	1.1250	1.0922	0.9929	1.3967	1.0276	1.1638	0

Table 9 Log Hedonic Distance Matrix

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	0	0.2505	0.5080	0.6212	0.9160	1.3046	0.7892
soy milk	0.2505	0	0.3379	0.4791	0.9679	1.1088	0.7109
rice milk	0.5080	0.3379	0	0.6376	1.0891	1.0327	0.7262
2% milk	0.6212	0.4791	0.6376	0	1.1105	1.1427	0.9482
1% milk	0.9160	0.9679	1.0891	1.1105	0	1.6581	0.7775
fat-free milk	1.3046	1.1088	1.0327	1.1427	1.6581	0	1.4205
whole milk	0.7892	0.7109	0.7262	0.9482	0.7775	1.4205	0

Table 10 Continuous Linear Hedonic Distance Matrix after Rescaling between 0 and 1

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	1	0.7929	0.6492	0.5763	0.6581	0.7207	0.4706
soy milk	0.7929	1	0.7250	0.6284	0.7403	0.8323	0.4780
rice milk	0.6492	0.7250	1	0.5626	0.6358	0.7362	0.5018
2% milk	0.5763	0.6284	0.5626	1	0.6314	0.5961	0.4172
1% milk	0.6581	0.7403	0.6358	0.6314	1	0.7080	0.4932
fat-free milk	0.7207	0.8323	0.7362	0.5961	0.7080	1	0.4621
whole milk	0.4706	0.4780	0.5018	0.4172	0.4932	0.4621	1

Table 11 Continuous Log Hedonic Distance Matrix after Rescaling between 0 and 1

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	1	0.7997	0.6631	0.6168	0.5219	0.4339	0.5589
soy milk	0.7997	1	0.7475	0.6761	0.5082	0.4742	0.5845
rice milk	0.6631	0.7475	1	0.6107	0.4787	0.4920	0.5793
2% milk	0.6168	0.6761	0.6107	1	0.4738	0.4667	0.5133
1% milk	0.5219	0.5082	0.4787	0.4738	1	0.3762	0.5626
fat-free milk	0.4339	0.4742	0.4920	0.4667	0.3762	1	0.4131
whole milk	0.5589	0.5845	0.5793	0.5133	0.5626	0.4131	1

Neighbors and Closeness

Practically, the concept or notion of neighbors borrows from the clustering paradigm[2]. In this section, we discuss the notion of neighbors and how the closeness index is obtained. A point's neighbors are those points that are considered similar to it. Let sim (p_i, p_j) be a similarity function that is normalized and captures the closeness between the pair of points p_i and p_j (Wu et al., 2013). The function *sim* could be one of the well-known distance metrics (Euclidean distance metrics) or it could even be non-metric (e.g., a distance/similarity function provided by a domain expert) (Guha et al., 2000). It is usually assumed that *sim* is valued between 0 and 1, with larger values indicating that the points are more similar. Given a threshold θ between 0 and 1, a pair of points p_i, p_j are defined to be neighbors if the following holds:

(18) $\text{sim}\ (p_i, p_j) \geq \theta$

In the above equation, θ is a user-defined parameter that can be used to control how close a pair of points must be in order to be considered neighbors. Thus, higher values of θ correspond to

[2] Clustering is the process of partitioning a set of patterns into cohesive groups or clusters. Such a process is carried out so that intra-cluster patterns are similar and inter-cluster patterns are dissimilar (Van Dongen, 2000). There exist many clustering algorithms. However, it is not clear what a "correct" clustering for that set is. Indeed, different algorithms may yield dramatically different outputs for the same input sets.

a higher threshold for the similarity between a pair of points before they are considered neighbors. If we assume that *sim* is equal to1 for identical points and equal to 0 for totally dissimilar points, assigning a value of 1 for θ constrains a point to be a neighbor to only other identical points. On the other hand, a value of 0 for θ permits any arbitrary pair of points to be neighbors. Depending on the desired closeness, an appropriate value of θ may be chosen by the user.

Concluding from the above argument, closeness and neighbors are related concepts in measuring the relative locations of two objects or points in a space. For many people, their first impression about closeness might be related to ecological or social communities. It is first introduced by Shimbel (1953). Closeness can be traced back to concepts appearing in sociology (Bavelas, 1948), geography (Haggett and Chorley, 1969) and transportation (Garrison and Marble, 1962; Ford and Fulkerson, 1962; Kansky, 1963). It is also a very basic concept in topological space in the area of Topology and related mathematical area. Intuitively, two sets can be interpreted as "close" if the one is close to another. In metric space, closeness can be naturally defined with the concept of "distance" which corresponding so much with the concept we used in this article. When we mention closeness, there is another concept called similarity which is very similar with closeness in some sense but defined in a different way. In recent studies about social and spatial networks, the concept of "closeness" and "similarity" are used. Corresponding to these two concepts, researchers developed two indexes to measure them. The basic definition of a similarity index is restricted to be a single number, which is a function of the pairwise comparisons of the values for each attribute for two samples (Johnston, 1976). In this work, he considered the "attribute space" or "universe of comparison" to be the complete set of attributes for which comparisons are possible. His research framework illustration is very helpful for demonstrating the practicability of hedonic attribute space. In general, similarity index, as indicated by Johnston, is characterized as the result of a two-step process defined on a pair of vectors. In the first step, an attribute similarity score is obtained for each attribute by comparing the attribute values observed in the pair of vectors. The result is a vector of attribute similarity scores, which is similar to the value added terms of all the attributes. Then the vectors of scores are combined in the second step to arrive at the similarity index. This step also can correspond to the calculation of hedonic distance based on the differences of all the attributes (equation 9 and equation 10). The operation in the first step was characterized as a function, G, defined on pairs of attribute values. The second operation was characterized as a function, F, defined on the vector of attribute similarity scores from the first step. Usually, F is a simple sum or weighted sum of the attribute similarity scores.

Based on the telecommunication work of Shimbel (1953) and Freeman (1977), closeness is developed as a measure of network accessibility flow potential. Basically, it measures a form of centrality based on shortest paths through the network. Modern implementations of these algorithms on spatial networks use a variety of distance metrics to define the shortest paths. Metrics may, for example, be Euclidean (minimizing the number of meters travelled along the network), angular (minimizing the cumulative angle turned along each route), topological (minimizing the number of nodes) or indeed based on collected data such as average journey times on individual links (Cooper, 2015). Since "similarity" has connotations of closeness in some sense and the hedonic distance matrix are constructed in the form of vectors, an attempt is made to relate closeness index to a vector space model with n dimensions, n being the total number of attributes (number of attributes the products contain). In this way, "closeness" can be objectified as "distance".

The nearest neighbor concept in aforementioned research field is, to a large extent, the same with it considered in hedonic matrix approach except for the elements that construct the matrix. Different from previous research, the nearest neighbor we calculated in this work is based on all the attributes from the hedonic regression which means the spaces are not arbitrarily chosen. Referring to the concepts of neighbors, similarity, and closeness, in this study, two products are nearest neighbors if they have the smallest value in the hedonic distance matrix and thus the continuous cross product closeness index is 1 between two products if they are nearest neighbors. Therefore, the nearest neighbor dummy matrix (as shown in table 12 and table 13) created based on table 10 and 11 are shown as follows.

Looking at each row of table 12, it is evident that the nearest neighbor (closest product) for each product. For example, in the first row, the smallest distance between almond milk and other products is associated with soy milk (valued 1), so it is the nearest neighbor of almond milk. In the same logic, for soy milk, rice milk, 2% milk, 1% milk, whole fat milk and fat-free milk, the closest products are fat-free milk, fat-free milk, 1% milk, soy milk, soy milk, and rice milk respectively. In the dummy matrix of log hedonic distance matrix, the nearest neighbor of almond milk, soy milk, rice milk 2% milk, 1% milk, fat-free milk and whole milk are soy milk, almond milk, soy milk, soy milk, whole milk, rice milk and soy milk respectively.

Table 12 Nearest Neighbor Dummy Matrix of Linear Hedonic Distance Matrix

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond	0	1	0	0	0	0	0
soy	0	0	0	0	0	1	0
rice	0	0	0	0	0	1	0
2% milk	0	0	0	0	1	0	0
1% milk	0	1	0	0	0	0	0
fat-free milk	0	1	0	0	0	0	0
whole milk	0	0	1	0	0	0	0

Table 13 Nearest Neighbor Dummy Matrix of Log Hedonic Distance Matrix

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond	0	1	0	0	0	0	0
soy	1	0	0	0	0	0	0
rice	0	1	0	0	0	0	0
2% milk	0	1	0	0	0	0	0
1% milk	0	0	0	0	0	0	1
fat-free milk	0	0	1	0	0	0	0
whole milk	0	1	0	0	0	0	0

From table 12 and 13 we have obtained the parameters-cross-product closeness index d_{ij}^{nn}. To calculate own closeness index x_i^c, according to equation 9 and 10, for each product, its own closeness index is one, because the distance of one product with itself is zero. Therefore, x_i^c are all ones for the seven products considered in this study. However, the list of such parameters cause singularity problem in estimating the demand system which leads to biased estimation of parameters, hence elasticities. Therefore, instead of using the own closeness index directly indicated by the formula, we have tried to use another method to measure the own closeness index. After exploring the similarity of using closeness and matrix distance between this study and some research methods from social networks, we find that the concepts and definitions of closeness centrality which is commonly used to measure the closeness of an object with other comparable objects in the social networks can be compared and applied to our study. According to the social network paradigm, in this study, we treat seven different products as seven nodes in the network and each node is connected with the other six nodes.

Closeness Centrality and Closeness Index

According to Wang and Tang (2014), two variants are defined in the context of an undirected, unweighted and connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1,3, \ldots, N\}$ denotes the set of $N \geq 2$ nodes and $\mathcal{E} \subset \{\{i, j\}\} : i, j \in \mathcal{V}, : i \neq j\}$ denotes the set of edges.

Another related concept need to mention is closeness centrality. It is a basic centrality measure that characterizes how centrally located a node is, within a network, based on its distances to all other nodes. Closeness centrality originally defined by Sabidussi (1966) is to assign higher scores to those who have shorter distances to all other nodes. Closeness centrality highlights the players who will be able to contact easily all other members of the network.

Before introducing the computation of classic "closeness index" being applied in our study, we first introduce some preliminaries. In our case, the graph is a directed, weighted and connected graph *G* which has *N* node $\mathcal{V}$ and *(N-1) N* edges in $\mathcal{E}$ which has circles. This is not a tree graph as the common network graph, the edges are all connected because we suppose they all have correlation between each other and they have distance from one to the other too. So for the seven products, there are 7 nodes and 42 edges. For each edge denoted as $\{i, j\} \in \mathcal{E}$, there is a corresponding edge $\{j, i\}$. For convenience, let $\mathcal{V}_{(i,j)}$ and $\mathcal{V}_{(ji,)}$ denote, respectively, the sets of nodes in these two connected components as illustrated in figure 8 and 9. $\mathcal{V}$ can be partitioned into $\{i\}$ and $\mathcal{V}_{(ij)} \forall j \in \mathcal{N}_i$, where $\mathcal{N}_i = \{j \in \mathcal{V}: \{i, j\} \in \mathcal{E}\}$ denotes the set of neighbors of node i.

The classic closeness often referred to simply as closeness C_i of the node $i \in \mathcal{V}$ indicated by Sabidussi (1966) can be represented as:

(19) $C_i \triangleq \frac{\mathrm{N}-1}{\sum_{j \in \mathcal{V}} d_{ij}}$,

where $d_{ij} = d_{ji}$ is the distance (i.e., length of the shortest paths) between nodes *i* and *j*, and the factor $N - 1$ is inserted so that $C_i \in (0,1]$. This is an oldest and most fundamental measures that have been widely used. It follows that, the larger the C_i, the closer node *i* is, on average, to all other nodes in the graph $\mathcal{G}$. Indeed, this measure has been applied to several areas, including epidemiology (Borgatti, 1995) social networks (Okamoto at al., 2008) and power systems (Wang, 2010; Nasiruzzaman et al., 2011). However, a limitation of classic closeness C_i indicated by Wang and Tang (2014) in equation 12 is that if node *i* is very far away from some node *j*, then even if node *i* is very close to the rest of the nodes, its C_i would be practically zero, signifying that node *i* has poor closeness. They suggest that in some applications, it is desirable to discount the the influence of those nodes that are very far away, thus preventing them from “skewing” the closeness of node *i*.

To solve this problem, they developed a new index named exponential closeness was developed,

(20) $C_i^E \triangleq \sum_{j \in \mathcal{V}, j \neq i} 2^{-d_{ij}}$

This is a measure that possesses desirable property that mitigates the effects of distant nodes. They also generalized the definition of C_i^E is generalized to be:

(21) $C_i^E \triangleq \sum_{\substack{j \in \mathcal{V}, \\ j \neq i}} \alpha^{-d_{ij}}$,

where $\alpha > 1$. This definition allows the base of the exponent to be a number other than 2. Considering the applicability of closeness centrality in our study to measure to measure own closeness index, we decided to use the closeness algorithm from social network study to acquire own closeness index. According to the study of Wang and Tang (2014), an algebraic relationship is established to expresses the classic closeness C_i as functions of some variables as presented in equation 12. The left two figures of figure 8 and 9 shows that each product is a node and connected with other nodes by a symmetric edge with the length of the edges being the Euclidean hedonic distance between this node and other nodes. The right two figures illustrate the closeness centrality of each node with other nodes, in which the larger C_i , the bigger node i is drawn. In additional to using classic closeness measurement to calculate own closeness index, we also used the exponential index to make comparisons. Table 14 shows the calculated classic closeness index and exponential closeness index of both linear and log forms.

Table 14 Classic Closeness Index and Exponential Closeness Index

	classic closeness index		exponential closeness index	
	Liner	Log	Linear	Log
almond milk	1.6812	1.3669	4.0433	3.7074
soy milk	2.0860	1.5564	4.3898	3.9351
rice milk	1.6569	1.3852	3.9946	3.7006
2% milk	1.2599	1.2148	3.5200	3.4447
1% milk	1.7306	0.9204	4.0631	2.8746
fat-free milk	1.8744	0.7825	4.2371	2.5007
whole milk	0.8826	1.1168	2.7467	3.2694

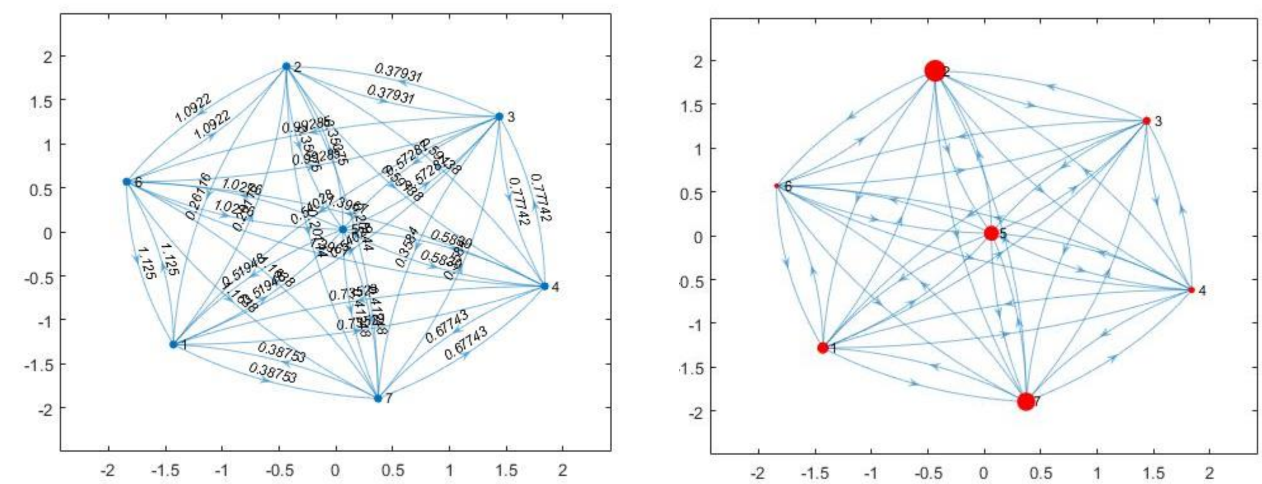


Figure 8 Classic Closeness Index Based on Linear Hedonic Distance of Seven Products

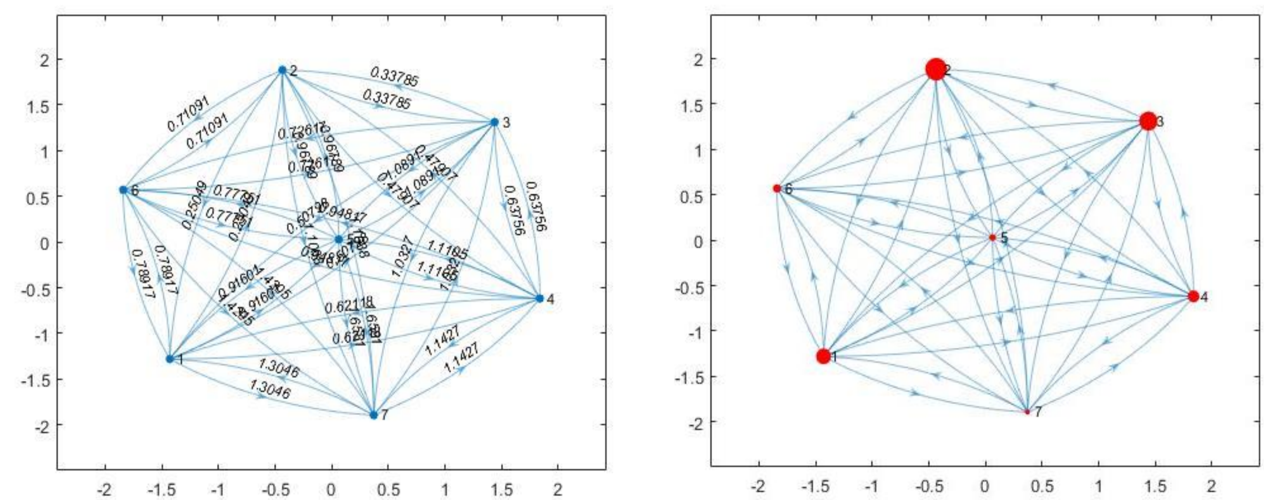


Figure 9 Classic Closeness Index Based on Log Hedonic Distance of Seven Products

Demand Systems

Demand system models are widely used in estimating demand relationships among wide range of food products. The Almost Ideal Demand System (AIDS) and Rotterdam model are two commonly used such demand systems. For almost three decades, the AIDS model developed by Deaton and Muellbauer (1980b) is one of the most widely used flexible demand specifications, which has been applied to examine consumer demand for varies agricultural commodities (Jabarin, 2005; Thompson, 2004; Richards et al., 1997; Mdafri, 1993; Heien and Pompelli, 1988; Blanciforti and Green, 1983).

One of the major advancements in demand system modeling was the development of the Rotterdam model by Theil (1965) and Barten (1964). Also these demand systems must satisfy a host of regularity conditions emanating from the demand theory (such as homogeneity and symmetry). These demand models allow for these conditions to be imposed or for the system to be estimated without restrictions to test the consistency of economic theory with the data. The Rotterdam model is specified in terms of first differences of the variables, thus making it a particularly attractive when time series of prices and income are nonstationary. For these reasons the Rotterdam model continues to be popular for purposes of demand analysis and testing of economic theory.

Barten (1993) introduced Barten's synthetic model (BSM), which includes the differential versions of the Almost Ideal Demand System (AIDS) model introduced by Deaton and Muellbauer (1980a, 1980b), the Rotterdam model introduced by Barten (1964) and Theil (1965), the NBR model introduced by Neves (1987), and the Dutch Central Bureau of Statistics (CBS) model introduced by Keller and van Driel (1985). The BSM model possesses a few characteristics that make it acclaimed in empirical research. These characteristics include linearity in parameters, functional form flexibility, ability to introduce dynamics, and potential to render variables stationary because of the necessity of the first-differencing process. In addition, Barten's differential demand system assists in identifying specific functional form that is best supported by the data. Following Matsuda (2005), the basic BSM is as follows:

(22) $w_i dlnq_i = (\beta_i + \lambda w_i) dln\mathrm{Q} + \sum_{j=1}^{n} [\gamma_{ij} - \mu w_i (\delta_{ij} - w_j)] dln\mathrm{p}_j$

where i =1,2,....,n and $\beta_i \equiv (1-\lambda)b_i + \lambda c_i$, and $\gamma_{ij} \equiv (1-\mu)s_{ij} + \mu r_{ij}$; $w_i \equiv (p_i q_i)/m$ denotes the expenditure share of good i, which determines the allocation of additional expenditure to the good; $dln\mathrm{Q} \equiv \sum_i w_i dln\mathrm{p}_i$ denotes the Divisia volume index. δ_{ij} denotes the Kronecker delta, which is equal to unity if i=j and zero otherwise; b_i are constant coefficients and $w_i \equiv b_i - w_i$. Depending on the restrictions we impose on coefficients μ and λ in equation (19), we recover the Rotterdam, the LA/AIDS, the CBS and the NBR models. (λ,μ) = (0,0) yields the Rotterdam model; (λ, μ) = (1,0) yields the CBS model; (λ, μ) = (0,1) gives rise to the NBR model; (λ, μ) = (1,1) yields the AIDS model.

To satisfy the theoretical properties associated with the demand theory, we assume following restrictions on parameters of BSM. Restrictions imposed are, adding-up:

(23) $\sum_{i=1}^{n} \beta_i + \lambda = 1$

(24) $\sum_{i=1}^{n} \gamma_{ij} = 0$

and homogeneity:

(25) $\sum_{j=1}^{n} \gamma_{ij} = 0$, where i=1,2,.....,n.

Slutsky symmetry condition is satisfied via the restriction:

(26) $\gamma_{ij} = \gamma_{ji}$ for i, j =1,2,......,n and $i \neq j$

Compensated price elasticity formula is expressed as follows:

(27) $e^{C}{}_{ij} = \frac{\gamma_{ij}}{w_i} - \mu(\delta_{ij} - w_j)$

where δ_{ij} is the Kronecker delta (δ_{ij} = 1 *if i* = *j* and δ_{ij} = 0 *if* $i \neq j$).

Now we show how compensated elasticity is derived. First we divide equation (19) by $w_i$to get the following:

(28) $d\ln q_i = \left(\frac{\beta_i}{w_i} + \lambda\right) d\ln Q + \left[\frac{\gamma_{ij}}{w_i} - \mu(\delta_{ij} - w_i)\right] d\ln p_j$

Now let us differentiate above equation with respect to $d\ln p_j$, to obtain the compensated price elasticity formula:

(29) $\frac{d\ln q_i}{d\ln p_j} = e^{C}{}_{ij} = \frac{\gamma_{ij}}{w_i} - \mu(\delta_{ij} - w_j)$

We recover the uncompensated price elasticities e_{ij}^{U} using the Slutsky derivative expressed in elasticity form as follows:

(30) $e_{ij}^{U} = e^{C}{}_{ij} - e_i w_j$

Next, compensated cross price elasticities were used to assess the symmetry conditions using following expression:

(31) $e^{C}{}_{ij} = \left(\frac{w_{ij}}{w_i}\right) e^{C}{}_{ji} + w_j(e_j - e_i)$

where w 's are budget shares of *i*th and *j*th good and, e_j and e_i are expenditure elasticities of *j*th and *i*th good respectively. Expenditure elasticity formula for Barten synthetic system is given as follows:

(32) $e_i = \frac{\beta_i}{w_i} + \lambda$

Now, let us show how to derive the expenditure elasticity.

We know that:

(33) $d\ln m = d\ln P + d\ln Q$

then we write above equation in terms of $d\ln Q$:

(34) $d\ln Q = d\ln m - d\ln P$

Applying above result from equation 18 into the Barten synthetic model equation gives us the following:

(35) $w_i d\ln q_i = (\beta_i + \lambda w_i)(d\ln m - d\ln P) + \sum_{j=1}^{n}[\gamma_{ij} - \mu w_i\,(\delta_{ij} - w_j)]d\ln p_j$

Simplifying equation further would result in the following:

(36) $w_i d\ln q_i = (\beta_i + \lambda w_i)d\ln m - (\beta_i + \lambda w_i)d\ln P + \sum_{j=1}^{n}[\gamma_{ij} - \mu w_i\,(\delta_{ij} - w_j)]d\ln p_j$

Dividing above equation through by w_i would give us the following expression:

(37) $d\ln q_i = \left(\frac{\beta_i}{w_i} + \lambda\right)d\ln m - \left(\frac{\beta_i}{w_i} + \lambda\right)d\ln P + \sum_{j=1}^{n}\left[\frac{\gamma_{ij}}{w_i} - \mu(\delta_{ij} - w_j)\right]d\ln \mathrm{p}_j$

Now, differentiating above equation with respect to $d\ln m$ would give us the formula for expenditure elasticity:

(38) $\frac{d\ln q_i}{d\ln m} \equiv e_i = \frac{\beta_i}{w_i} + \lambda$

After the re-parameterization of the barten synthetic model which include:

(39) $\gamma_{ii}=\mathrm{c}_h d_{ij}^{h} + \mathrm{c}_{nn} d_{ij}^{nn}$

(40) $\gamma_{ij} = \alpha_0 + \alpha_1 x_i^s + \alpha_2 x_i^c$

Where x_i^s refers to the market share and x_i^c is the closeness index of product i. Putting these coefficients back in the original model, we get the following Hedonic Metric augmentation to the Barten Synthetic model (HM-BSM) which allows us to implement differentiated quality attributes in demand elasticity estimations.

(41) $w_i d\ln q_i = (\beta_i + \lambda w_i)d\ln \mathrm{Q} + (\alpha_0 + \alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_i(\delta_{ii} - w_i))d\ln p_i +$

$$\sum_{\mathrm{i}\neq\mathrm{j}}\left[\mathrm{c_h d_{ij}^h + c_{nn} d_{ij}^{nn}} - \mu \mathrm{w_i}(\delta_{\mathrm{ij}} - \mathrm{w_j})\right]\mathrm{dlnp_j}$$

In this form, the Hicksian own-price elasticity is:

(42) $\frac{d\ln q_i}{d\ln p_i} = e^{C}{}_{ii} = \frac{\alpha_0+\alpha_1 x_i^S+\alpha_2 x_i^C}{w_i} - \mu(\delta_{ii} - w_i)$

the Hicksian cross-price elasticities can be calculated as

(43) $\frac{d\ln q_i}{d\ln p_j} = e^{C}{}_{ij} \frac{\mathrm{c}_h d_{ij}^h+\mathrm{c}_{nn} d_{ij}^{nn}}{w_i} - \mu(\delta_{ij} - w_j)$

The expenditure elasticities are calculated in the original form from the original model.

(44) $\frac{d\ln q_i}{d\ln m} \equiv e_i = \frac{\beta_i}{w_i} + \lambda$

The Marshallian elasticities are recovered using the Slutsky equation in elasticity form

(45) $e_{ij}^U = e^{C}{}_{ij} - e_i w_j$, as suggested by Barnett and Serlitis (2008).

Therefore, the Marshallian own-price elasticities are:

(46) $e_{ij}^U = \frac{\alpha_0+\alpha_1 x_i^S+\alpha_2 x_i^C}{w_i} - \mu(\delta_{ii} - w_i) - e_i w_i,$

where e_i is the expenditure elasticities which equal to $\frac{\beta_i}{w_i} + \lambda$. So,

(47) $e_{ij}^U = \frac{\alpha_0+\alpha_1 x_i^S+\alpha_2 x_i^C}{w_i} - \mu(\delta_{ii} - w_i) - \beta_i + \lambda w_i.$

When estimating a complete system of demand equations, a common problem is that there are many parameters relative to the number of observations available to estimate. This method (Distance Metric augmented Barten synthetic model DM-BSM) offers solution to price endogeneity issue common with estimating complete demand system as well as reduce the number of parameters available to estimate.

CHAPTER III

DATA

Nielsen Homescan[3] consumer panel data 2004-2015 is used in this analysis. Consumer level data are gathered by tracking households' purchase behavior. Each purchase of a household is recorded using a scanner, and subsequently uploaded to a database. From this large panel of data, weekly purchases of 3,000 households were extracted, who regularly participated in the survey from 2004 to 2015. This database is used to obtain price, quantity, and expenditure data of three dairy alternative beverages (almond milk, soy milk andrice milk) and four types of conventional milk (2% milk, 1% milk, fat-free milk, and whole milk).

The quantity purchased, total price paid and other purchase information in each trip of 3,000 households are used to derive the quantity and unit monthly average prices. The final data set is aggregated to UPC level which captures not only enough variation of nutritional variables but also consumer purchase information.

Data for Estimating Hedonic Pricing Models

As shown in tables 15-21, the descriptive statistics of variables used in this analysis are listed. Table 22 and 23 are correlation matrix of log of prices of seven products. The number of variables varies as for different products, their package sizes and multi-pack are different. Price variable, as described, is a unit value. Because we also estimate the log regression, log prices are added in the tables. For nutritional variables, the values are based on 8 oz. (1 cup) for each

[3] Calculated (or derived) based on data from The Nielsen Company (U.S.), LLC and marketing database provided by the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from The Nielsen data are those of the researchers and do not reflect the views of the Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

product. The unit of variables including fat, fiber, and protein are grams and that of vitamin A, Vitamin D and calcium are percent value. It can be shown that the average prices of dairy alternative beverages are generally higher than conventional fluid milk products.

Three types of dairy alternative beverages (almond milk, soy milk and rice milk) and the four most common types of milk products (whole milk, 1% milk, 2% milk and fat free milk (or skim milk) are included in this work and monthly average price variable is acquired as follows. Monthly average price is the “unit price paid” as shown in the last box below.

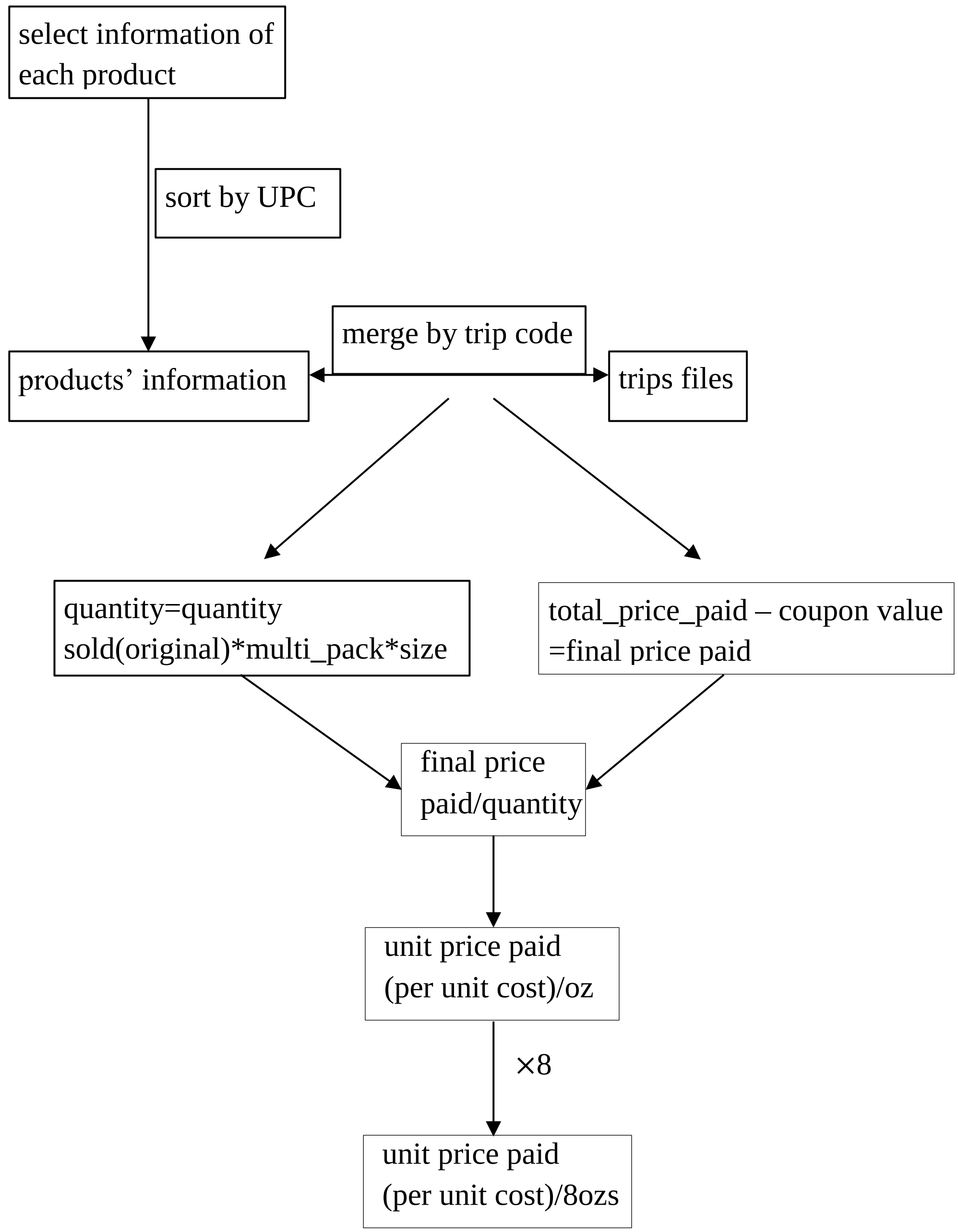


Figure 10 Steps to Calculate Variable Unit Price Paid

First, we obtain each product's information from products files; and then we merge the information with trips files to acquire the dataset, which include variables of quantities sold, total price paid by consumers, coupon value, deal_flag_uc, multi_pack, product's package size and size unit[4]. As figure 10 shows, the unit price paid (per unit cost) is calculated by first dividing the final price paid by the quantity variable created above. Final_price_paid is calculated by subtracting the value of variables "coupon value" from the value of "total_price_paid". Then, we average the unit prices paid in each month in each year to get the monthly average price per oz. and multiply by 8 to get monthly average price per 8oz. (unit monthly average price).

Obtaining data on nutritional information of dairy milk alternative beverages is one of the biggest concerns in estimating the hedonic pricing models due to unavailability of proper database pertains to such information. In this work, the nutritional variables are obtained from search for the product label of different products. The USDA Nutritional Database and IRI nutrition information databases are widely used references for most food composition studies. However, even though USDA Nutritional Database includes 50 different categories for varieties of milk products, most of them are conventional milk, with little available information on dairy alternative beverages. Although IRI database has nutritional information of dairy alternative beverages, their database uses different UPC system with Nielsen and the nutritional information recorded are not based on uniform unit. Therefore, the only way to obtain the nutritional information is gathering the nutritional information directly from the products' label by individual visual observations of beverage packages. The final dataset reflects the same set of qualitative information that consumers have about these products based on their labels. After we

[4] size1_units is the unit of measure. For example, "size1_amount" might be "16.0", and " size1_units" might be "OZ."; coupon is total discount for amount due to coupon; deal_flag_uc is presence of a deal (1=deal, 0=no deal).

obtain the detailed product characteristics data, we merge this with Nielsen based on the barcode, or Universal Product Code (UPC) to construct the complete dataset for estimating hedonic pricing models.

Besides the price variable and nutritional variables necessary in the hedonic model estimation, we attempt to find out other variables that might exert impact on the prices of those products. The first group of such variables includes package size and multi-pack. We also consider that deal/coupon available when consumers purchase the products should affect prices. Therefore, we select two variables which are deal_flag_uc and coupon_value. "deal_flag_uc" is a dummy variable which indicates if the panelist received a deal. Also if the panelist used a coupon, they enter the amount discounted. If coupon_value and deal_flag_uc are both zero, there is no deal on the purchase. In addition, in order to take into consideration of the time effects on the prices, we add yearly dummies. Since the format of original data file in which households' purchase information is recorded by their trip date, it is very common that the purchase may happen many times each month or no purchase activities within a month. That's one reason we aggregate the data into UPC level. Another very important variable considered to affect price is brand. Therefore a brand dummy is added which equals 1 if it is a store brand and equals zero if it is a national brand.

Data for Estimating the Demand System

The process begins with the original datasets which are used to estimate the hedonic regressions. We use "expoz" to denote the variable final price paid (the amount of ounces consumed) which is calculated by expoz=total_price_paid –coupon value. After this, we create the average price for the products based on the month and year instead of UPC, month and year. The reason why we need to create the average price again is because in the original dataset the prices we used previously is monthly average prices for UPC level average but now we need monthly average prices for each product (total seven products) and we also create the average quantity because of the same reason.

Then I sum “expoz” to create a variable “sumexpoz” based on UPC, month and year and delete the duplicated observations. The reason why we add UPC sorting level is because this enables us to do future research on UPC level demand. From these two steps, we acquired the dataset which include the non-repeated UPC level products purchased on each month. This aims for creating the w_i (budget share) in the final step. After this, we create the variable “asumexpoz” which is the sum of “sumpoz” based on month and year for each UPC-level product. From this step, we get the aggregate product level expenditures every month every year. After this, the dataset for each product include observations that have the same expenditure every month for every year. Therefore, deleted the repeated observations again based on variables “asumexpoz”, month and year. Then the dataset is monthly expenditure observations from year 2004 to 2015 for each product. Then we merge the seven datasets in order to calculate the budget share. After merging the dataset, we turn to construct another new variable “asumtotal” by summing the variable asumpoz based on month and year. Then we get the total expenditure of seven products each month each year. Now we use “asumpoz” divided by “asumtotal” to acquire the products’ budget share. Because we have seven products with average budget shares from year 2004 to 2015, the final data set contain 145 observations for each product and the whole system has 1015 (7 products and each has 145 observations) observations in the whole demand system.

Again, from this dataset, we are able to calculate the budget share of each product using the total monthly expenditure of each product divided by monthly total expenditure of seven products. After combing all the variables to construct final dataset, we construct the variables used in the demand system estimation.

Average Expenditure and Prices

Figure 11 shows the average expenditure of seven products from 2004 to 2015. Accordingly, during this time period the expenditures for the four types of regular milk products did not vary much. However, during year 2007 to year 2009, all types of milk alternative beverages and conventional milk products show growth in expenditure with rice milk being the most significant category. Furthermore, we can observe that in the year 2007 and 2008, rice milk has experienced a fast growth. Also for almond and soy milk, their consumption also increased in the same period even though not as much as rice milk.

A price checking on FreshDirect.com shows that half a gallon of soy milk costs \$3.99 to\$4.29, compared to \$2.39 to \$2.49 for half a gallon of regular fat-free and reduced-fat milk. Accordingly, from figure 12 we can see that the unit price of soy milk in 2004 is almost three times the other types of regular milk products. Its price decreases a lot from year 2004 to the mid of year 2006 and then it has a slight increase during 2007-2008 and decrease to almost the same price with dairy milk from year 2011 to 2015. According to figure 11, the rising expenditure of soy milk and almond milk between year 2008 and 2009 is caused mostly by the high price of those two dairy milk alternative beverages.

Table 15 Summary Statistics: Almond Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	1,591	0.6369	0.3020	0	3.4171
$\ln(P_1)$	ln_price	1,591	-0.5225	0.3566	-2.9301	1.2288
x_{Kcal}	calories	1,591	77.4778	35.5234	30	200
x_{fat} (g)	fat	1,591	3.0439	0.6819	0	6.8571
x_{fiber} (g)	fiber	1,591	1.4773	1.5467	0	5
$x_{protein}$ (g)	protein	1,591	1.3351	0.9534	0.6667	6.1
x_{VA} (%)	Vitamin_A	1,591	14.2548	9.7966	0	30
x_{cal} (%)	calcium	1,591	38.0099	11.4905	4	60
x_{VD} (%)	VD	1,591	22.9395	7.6080	0	30
D_{deals}	deal_flag_uc	1,591	0.2476	0.4318		
D_{pkge_size1}	package_size1	1,591	0.0365	0.1875		
D_{pkge_size2}	package_size2	1,591	0.0038	0.0613		
D_{pkge_size3}	package_size3	1,591	0.0239	0.1527		
D_{pkge_size4}	package_size4	1,591	0.0050	0.0708		
D_{pkge_size5}	package_size5	1,591	0.0050	0.0708		
D_{pkge_size6}	package_size6	1,591	0.0383	0.1921		
D_{pkge_size7}	package_size7	1,591	0.7454	0.4357		
D_{pkge_size8}	package_size8	1,591	0.0302	0.1711		
D_{pkge_size9}	package_size9	1,591	0.1119	0.3153		
D_{multi2}	multi_2	1,591	0.0019	0.0434		
D_{multi4}	multi_4	1,591	0.0101	0.0998		
D_{multi6}	multi_6	1,591	0.0082	0.0901		
D_{multi7}	multi_7	1,591	0.0094	0.0967		
D_{multi8}	multi_8	1,591	0.0044	0.0662		
D_{coupon}	coupon	1,591	0.0830	0.2759		
$D_{year2014}$	year_14	1,591	0.1157	0.3199		
$D_{year2013}$	year_13	1,591	0.0924	0.2897		

Table 15 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
D_{year2012}	year_12	1,591	0.1025	0.3033		
D_{year2011}	year_11	1,591	0.0949	0.2932		
D_{year2010}	year_10	1,591	0.0874	0.2825		
D_{year2009}	year_09	1,591	0.0767	0.2662		
D_{year2008}	year_08	1,591	0.0685	0.2527		
D_{year2007}	year_07	1,591	0.0572	0.2323		
D_{year2006}	year_06	1,591	0.0459	0.2093		
D_{year2005}	year_05	1,591	0.0264	0.1604		
D_{year2004}	year_04	1,591	0.0308	0.1728		

Table 16 Summary Statistics: Soy Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	10,904	0.4183	0.2479	0	2.3855
$\ln(P_1)$	ln_price	10,873	-0.9638	0.3849	-4.1197	0.8694
x_{Kcal}	calories	10,010	105.5549	30.2654	50	193.3333
x_{fat} (g)	fat	10,010	3.3218	1.0906	0	6
x_{fiber} (g)	fiber	10,010	1.4017	0.8109	0	5
x_{protein} (g)	protein	10,010	6.6847	1.4877	1	16
x_{VA} (%)	Vitamin_A	10,010	10.5025	5.8497	0	150
x_{cal} (%)	calcium	10,010	32.6198	8.2041	0	60
x_{VD} (%)	VD	10,010	26.7098	9.5451	0	50
D_{deals}	deal_flag_uc	10,904	0.2063	0.4047		
D_{brands}	brands_s	10,904	0.2686	0.4433		
$D_{\text{pkge\_size1}}$	package_size1	10,904	0.0052	0.0721		
$D_{\text{pkge\_size2}}$	package_size2	10,904	0.0028	0.0524		
$D_{\text{pkge\_size3}}$	package_size3	10,904	0.0145	0.1195		
$D_{\text{pkge\_size4}}$	package_size4	10,904	0.0001	0.0096		
$D_{\text{pkge\_size5}}$	package_size5	10,904	0.0103	0.1008		
$D_{\text{pkge\_size6}}$	package_size6	10,904	0.0055	0.0740		
$D_{\text{pkge\_size7}}$	package_size7	10,904	0.0038	0.0612		
$D_{\text{pkge\_size8}}$	package_size8	10,904	0.0177	0.1319		

Table 16 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
$D_{\text{pkge\_size9}}$	package_size9	10,904	0.0024	0.0488		
$D_{\text{pkge\_size10}}$	package_size10	10,904	0.0920	0.2890		
$D_{\text{pkge\_size11}}$	package_size11	10,904	0.0027	0.0515		
$D_{\text{pkge\_size12}}$	package_size12	10,904	0.0006	0.0253		
$D_{\text{pkge\_size13}}$	package_size13	10,904	0.8400	0.3667		
$\mathrm{D}_{\text{pkge\_size14}}$	package_size14	10,904	0.0026	0.0506		
$\mathrm{D}_{\text{multi2}}$	multi_2	10,904	0.0205	0.1415		
$\mathrm{D}_{\text{multi3}}$	multi_3	10,904	0.0288	0.1672		
$\mathrm{D}_{\text{multi4}}$	multi_4	10,904	0.0003	0.0166		
$\mathrm{D}_{\text{multi6}}$	multi_6	10,904	0.0049	0.0696		
$\mathrm{D}_{\text{coupon}}$	coupon	10,904	0.0676	0.2511		
$\mathrm{D}_{\text{year2014}}$	year_14	10,904	0.1109	0.3140		
$\mathrm{D}_{\text{year2013}}$	year_13	10,904	0.1178	0.3224		
D_{year2012}	year_12	10904	0.131328	0.3378		
$\mathrm{D}_{\text{year2011}}$	year_11	10904	0.136189	0.3430		
$\mathrm{D}_{\text{year2010}}$	year_10	10904	0.136922	0.3438		
$\mathrm{D}_{\text{year2009}}$	year_09	10904	0.013756	0.1165		
$\mathrm{D}_{\text{year2008}}$	year_08	10904	0.005778	0.0758		
$\mathrm{D}_{\text{year2007}}$	year_07	10904	0.141416	0.3485		
$\mathrm{D}_{\text{year2006}}$	year_06	10904	0.110327	0.3133		
$\mathrm{D}_{\text{year2005}}$	year_05	10904	0.002843	0.0533		
$\mathrm{D}_{\text{year2004}}$	year_04	10904	0.001834	0.0428		

Table 17 Summary Statistics: Rice Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	839	0.4634	0.2557	0.0794	1.9950
$\ln(P_1)$	ln_price	839	-0.8698	0.4206	-2.5336	0.6906
x_{Kcal}	calories	838	136.6468	21.7658	90	200
x_{fat} (g)	fat	838	2.7584	0.5480	2	4
x_{fiber} (g)	fiber	838	0.2243	0.9714	0	5
x_{protein} (g)	protein	838	1.9499	1.2023	0	5
x_{VA} (%)	Vitamin_A	838	5.5519	3.9113	0	10
x_{cal} (%)	calcium	839	31.2420	3.0477	0	35
x_{VD} (%)	VD	838	13.9236	12.4797	0	26
D_{deals}	deal_flag_uc	839	0.1502	0.3575		
D_{brands}	brands_s	839	0.1025	0.3035		
$D_{\mathrm{pkge\_size1}}$	package_size1	839	0.0024	0.0488		
$D_{\mathrm{pkge\_size2}}$	package_size2	839	0.0346	0.1828		
$D_{\mathrm{pkge\_size3}}$	package_size3	839	0.0381	0.1917		
$D_{\mathrm{pkge\_size4}}$	package_size4	839	0.1120	0.3156		
$D_{\mathrm{pkge\_size5}}$	package_size5	839	0.1836	0.3873		
$D_{\mathrm{pkge\_size6}}$	package_size6	839	0.0751	0.2637		
$D_{\mathrm{pkge\_size7}}$	package_size7	839	0.5304	0.4994		
D_{multi2}	multi_2	839	0.0358	0.1858		
D_{coupon}	coupon	839	0.0298	0.1701		
D_{year2014}	year_14	839	0.0906	0.2872		
D_{year2013}	year_13	839	0.0822	0.2749		
D_{year2012}	year_12	839	0.1025	0.3035		
D_{year2011}	year_11	839	0.0918	0.2889		
D_{year2010}	year_10	839	0.0834	0.2767		
D_{year2009}	year_09	839	0.0930	0.2906		
D_{year2008}	year_08	839	0.0846	0.2785		
D_{year2007}	year_07	839	0.0810	0.2731		

Table 17 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
D_{year2006}	year_06	839	0.0667	0.2497		
D_{year2005}	year_05	839	0.0727	0.2598		
D_{year2004}	year_04	839	0.0727	0.2598		

Table 18 Summary Statistics: Whole Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	100,390	0.3119	0.1773	0	9.99
$\ln(P_1)$	ln_price	100,359	-1.2669	0.4377	-6.2792	2.3016
x_{Kcal}	calories	100,390	149.7963	4.3359	80	160
x_{fat} (g)	fat	100,390	7.9789	0.3079	5	9
x_{protein} (g)	protein	100,390	8.9530	0.9055	4.3333	11
x_{VA} (%)	vitamin_A	100,390	6.0639	0.7108	4	15
x_{cal} (%)	calcium	100,359	32.3247	2.4979	20.8333	60
x_{VD} (%)	VD	100,390	24.9657	0.9263	0	26
D_{deals}	deal_flag_uc	100,390	0.1150	0.3191		
D_{brands}	brands_s	100,390	0.4279	0.4948		
$D_{\text{pkge\_size1}}$	package_size1	100,390	0.0000	0.0063		
$D_{\text{pkge\_size2}}$	package_size2	100,359	0.0051	0.0710		
$D_{\text{pkge\_size3}}$	package_size3	100,390	0.0014	0.0375		
$D_{\text{pkge\_size4}}$	package_size4	100,390	0.0016	0.0398		
$D_{\text{pkge\_size5}}$	package_size5	100,390	0.0063	0.0792		
$D_{\text{pkge\_size6}}$	package_size6	100,390	0.0048	0.0691		
$D_{\text{pkge\_size7}}$	package_size7	100,390	0.0717	0.2580		
$D_{\text{pkge\_size8}}$	package_size8	100,359	0.0011	0.0335		
$D_{\text{pkge\_size9}}$	package_size9	100,390	0.0001	0.0114		
$D_{\text{pkge\_size10}}$	package_size10	100,390	0.1575	0.3643		
$D_{\text{pkge\_size11}}$	package_size11	100,390	0.0006	0.0242		
$D_{\text{pkge\_size12}}$	package_size12	100,390	0.0001	0.0109		
$D_{\text{pkge\_size13}}$	package_size13	100,359	0.0009	0.0303		
$D_{\text{pkge\_size14}}$	package_size14	100,390	0.3370	0.4727		

Table 18 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
$D_{\text{pkge\_size15}}$	package_size15	100,390	0.0045	0.0668		
D_{multi2}	multi_2	100,390	0.0038	0.0615		
D_{multi3}	multi_3	100,390	0.0007	0.0266		
D_{multi5}	multi_5	100,359	0.0003	0.0173		
D_{multi6}	multi_6	100,390	0.0000	0.0055		
D_{year2014}	year_14	100,390	0.0814	0.2735		
D_{year2013}	year_13	100,390	0.0826	0.2753		
D_{year2012}	year_12	100,390	0.0840	0.2774		
D_{year2011}	year_11	100,390	0.1617	0.3682		
D_{year2010}	year_10	100,390	0.0891	0.2849		
D_{year2009}	year_09	100,390	0.0785	0.2690		
D_{year2008}	year_08	100,359	0.1086	0.3111		
D_{year2007}	year_07	100,390	0.0145	0.1194		
D_{year2006}	year_06	100,390	0.0757	0.2645		
D_{year2005}	year_05	100,390	0.0768	0.2663		
D_{year2004}	year_04	100,390	0.0768	0.2663		

Table 19 Summary Statistics: 1% Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	49,422	0.2399	0.1031	0	2.4514
$\ln(P_1)$	ln_price	49,416	-1.4991	0.3698	-6.2792	0.8967
x_{Kcal}	calories	49,422	110.7803	4.9886	100	170
x_{fat} (g)	fat	49,422	2.6037	0.5089	2	8
x_{protein} (g)	vitamin_A	49,422	10.1064	0.7380	6	15
x_{VA} (%)	calcium	49,422	30.2281	1.8442	25	45
x_{cal} (%)	protein	49,416	8.0584	0.3556	8	11
x_{VD} (%)	VD	49,422	25.0017	1.3388	0	35
D_{deals}	deal_flag_uc	49,422	0.1726	0.3779		
D_{brands}	brands_s	49,422	0.5279	0.4992		
$D_{\text{pkge\_size1}}$	package_size1	49,422	0.0012	0.0342		

Table 19 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
$D_{\text{pkge\_size2}}$	package_size2	49,416	0.0005	0.0234		
$D_{\text{pkge\_size3}}$	package_size3	49,422	0.0001	0.0110		
$D_{\text{pkge\_size4}}$	package_size4	49,422	0.0009	0.0291		
$D_{\text{pkge\_size5}}$	package_size5	49,422	0.0003	0.0174		
$D_{\text{pkge\_size6}}$	package_size6	49,422	0.0022	0.0473		
$D_{\text{pkge\_size7}}$	package_size7	49,416	0.0791	0.2699		
$D_{\text{pkge\_size8}}$	package_size8	49,422	0.0022	0.0469		
$D_{\text{pkge\_size9}}$	package_size9	49,422	0.3813	0.4857		
$D_{\text{pkge\_size10}}$	package_size10	49,422	0.0001	0.0090		
$D_{\text{pkge\_size11}}$	package_size11	49,422	0.0040	0.0633		
$\mathrm{D}_{\text{pkge\_size12}}$	package_size12	49,416	0.0026	0.0510		
$D_{\text{pkge\_size13}}$	package_size13	49,422	0.5238	0.4994		
$D_{\text{pkge\_size14}}$	package_size14	49,422	0.0006	0.0242		
D_{multi2}	multi_2	49,422	0.0065	0.0806		
D_{multi3}	multi_3	49,422	0.0051	0.0715		
D_{multi6}	multi_6	49,416	0.0001	0.0119		
D_{multi7}	multi_7	49,422	0.0004	0.0201		
$\mathrm{D}_{\text{coupon}}$	coupon	49,422	0.0401	0.1962		
D_{year2014}	year_14	49,422	0.0507	0.2194		
D_{year2013}	year_13	49,422	0.0519	0.2218		
D_{year2012}	year_12	49,416	0.0534	0.2249		
D_{year2011}	year_11	49,422	0.1227	0.3281		
D_{year2010}	year_10	49,422	0.0658	0.2479		
D_{year2009}	year_09	49,422	0.0018	0.0424		
D_{year2008}	year_08	49,422	0.1305	0.3368		
D_{year2007}	year_07	49,416	0.0734	0.2607		
D_{year2006}	year_06	49,422	0.1679	0.3738		
D_{year2005}	year_05	49,422	0.1159	0.3201		
D_{year2004}	year_04	49,416	0.1155	0.3197		

Table 20 Summary Statistics: 2% Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	117,566	0.2911	0.1717	0	29.9950
$\ln(P_1)$	ln_price	117,536	-1.3313	0.4257	-7.3778	3.4010
x_{Kcal}	calories	117,566	129.7193	8.8784	80	190
x_{fat} (g)	fat	117,566	4.9600	0.5683	0	9
x_{protein} (g)	protein	117,536	8.1132	0.4813	7	10
x_{VA} (%)	vitamin_A	117,566	9.8992	0.6224	0	15
x_{cal} (%)	calcium	117,566	29.9992	1.2736	25	35
x_{VD} (%)	VD	117,536	22.9741	6.7407	0	25
D_{deals}	deal_flag_uc	117,566	0.1298	0.3361		
D_{brands}	brands_s	117,566	0.4268	0.4946		
$D_{\text{pkge\_size1}}$	package_size1	117,536	0.0000	0.0029		
$D_{\text{pkge\_size2}}$	package_size2	117,566	0.0082	0.0899		
$D_{\text{pkge\_size3}}$	package_size3	117,566	0.0009	0.0297		
$D_{\text{pkge\_size4}}$	package_size4	117,536	0.0015	0.0389		
$D_{\text{pkge\_size5}}$	package_size5	117,566	0.0014	0.0371		
$D_{\text{pkge\_size6}}$	package_size6	117,566	0.0002	0.0130		
$D_{\text{pkge\_size7}}$	package_size7	117,536	0.0067	0.0816		
$D_{\text{pkge\_size8}}$	package_size8	117,566	0.0024	0.0485		
$D_{\text{pkge\_size9}}$	package_size9	117,566	0.0490	0.2158		
$D_{\text{pkge\_size10}}$	package_size10	117,536	0.0005	0.0220		
$D_{\text{pkge\_size11}}$	package_size11	117,566	0.0001	0.0097		
$D_{\text{pkge\_size12}}$	package_size12	117,566	0.1357	0.3424		
$D_{\text{pkge\_size13}}$	package_size13	117,536	0.0006	0.0253		
$D_{\text{pkge\_size14}}$	package_size14	117,566	0.0001	0.0117		
$D_{\text{pkge\_size15}}$	package_size15	117,566	0.0001	0.0109		
$D_{\text{pkge\_size16}}$	package_size16	117,536	0.3610	0.4803		
$D_{\text{pkge\_size17}}$	package_size17	117,566	0.0001	0.0117		
$D_{\text{pkge\_size18}}$	package_size18	117,566	0.0045	0.0669		
$D_{\text{pkge\_size19}}$	package_size19	117,536	0.0013	0.0353		

Table 20 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
D_{multi2}	multi_2	117,566	0.0091	0.0952		
D_{multi3}	multi_3	117,566	0.0015	0.0381		
D_{multi6}	multi_6	117,536	0.0010	0.0318		
D_{multi7}	multi_7	117,566	0.0000	0.0029		
D_{coupon}	coupon	117,566	0.0280	0.1650		
D_{year2014}	year_14	117,536	0.0789	0.2695		
D_{year2013}	year_13	117,566	0.0804	0.2718		
D_{year2012}	year_12	117,566	0.0826	0.2753		
D_{year2011}	year_11	117,536	0.0825	0.2752		
D_{year2010}	year_10	117,566	0.0878	0.2830		
D_{year2009}	year_09	117,566	0.0895	0.2854		
D_{year2008}	year_08	117,536	0.0931	0.2905		
D_{year2007}	year_07	117,566	0.0957	0.2942		
D_{year2006}	year_06	117,566	0.0805	0.2721		
D_{year2005}	year_05	117,536	0.0789	0.2696		
D_{year2004}	year_04	117,566	0.0819	0.2741		

Table 21 Summary Statistics: Fat-free Milk

Variable	Label	N	Mean	Std Dev	Min	Max
P_1	price	88,259	0.2725	0.1486	0	7.9900
$\ln(P_1)$	ln_price	88,247	-1.3914	0.4119	-7.3778	2.0782
x_{Kcal}	calories	88,259	84.1034	5.6029	60	130
x_{fat} (g)	fat	88,259	0.0645	0.5950	0	8
x_{protein} (g)	protein	88,259	8.2004	0.4014	8	11
x_{VA} (%)	vitamin_A	88,259	9.7566	1.4145	0	15
x_{cal} (%)	calcium	88,259	30.2431	2.5912	25	60
x_{VD} (%)	VD	88,259	24.5033	3.4929	0	30
D_{deals}	deal_flag_uc	88,259	0.1527	0.3597		
D_{brands}	brands_s	88,247	0.4645	0.4987		
$D_{\mathrm{pkge\_size1}}$	package_size1	88,259	0.0042	0.0646		

Table 21 Continued

Variable	Label	N	Mean	Std Dev	Min	Max
$D_{\text{pkge\_size2}}$	package_size2	88,259	0.0010	0.0319		
$D_{\text{pkge\_size3}}$	package_size3	88,259	0.0002	0.0139		
$D_{\text{pkge\_size4}}$	package_size4	88,259	0.0000	0.0067		
$D_{\text{pkge\_size5}}$	package_size5	88,259	0.0049	0.0698		
$D_{\text{pkge\_size6}}$	package_size6	88,259	0.0005	0.0228		
$D_{\text{pkge\_size7}}$	package_size7	88,259	0.0149	0.1211		
$D_{\text{pkge\_size8}}$	package_size8	88,247	0.1265	0.3324		
$D_{\text{pkge\_size9}}$	package_size9	88,259	0.0002	0.0139		
$D_{\text{pkge\_size10}}$	package_size10	88,259	0.0003	0.0165		
$D_{\text{pkge\_size11}}$	package_size11	88,259	0.0003	0.0165		
$D_{\text{pkge\_size12}}$	package_size12	88,259	0.0001	0.0106		
$D_{\text{pkge\_size13}}$	package_size13	88,259	0.4165	0.4930		
$D_{\text{pkge\_size14}}$	package_size14	88,259	0.0002	0.0135		
$D_{\text{pkge\_size15}}$	package_size15	88,259	0.0065	0.0804		
$D_{\text{pkge\_size16}}$	package_size16	88,247	0.0014	0.0379		
$\mathrm{D}_{\text{multi2}}$	multi_2	88,259	0.0085	0.0917		
D_{coupon}	coupon	88,259	0.0350	0.1838		
D_{year2014}	year_14	88,259	0.0856	0.2798		
D_{year2013}	year_13	88,259	0.0896	0.2855		
D_{year2012}	year_12	88,259	0.0864	0.2809		
D_{year2011}	year_11	88,259	0.1027	0.3036		
D_{year2010}	year_10	88,259	0.0792	0.2701		
D_{year2009}	year_09	88,247	0.0267	0.1612		
D_{year2008}	year_08	88,259	0.1055	0.3073		
D_{year2007}	year_07	88,259	0.0587	0.2351		
D_{year2006}	year_06	88,259	0.0922	0.2893		
D_{year2005}	year_05	88,259	0.0872	0.2822		
D_{year2004}	year_04	88,259	0.0924	0.2895		

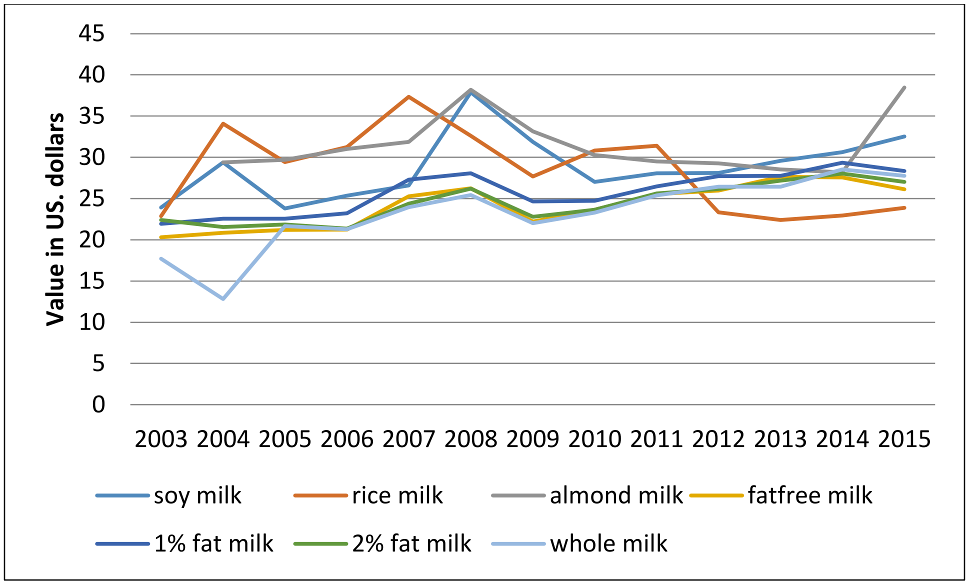


Figure 11 Average Expenditure of Dairy Alternative Beverages and Dairy Milk December 2003-December 2015

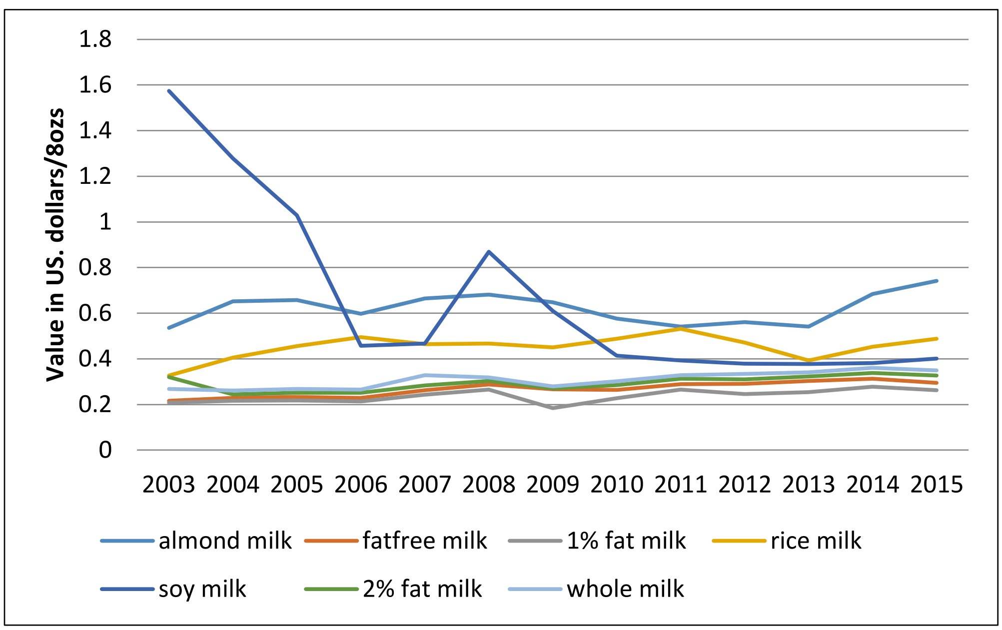


Figure 12 Average Price of Dairy Alternative Beverages and Dairy Milk December 2003-December 2015

Table 22 Summary Statistics of Log Prices of Seven Products

Variable	N	Mean	Std Dev	Sum	Min	Max
$\ln(p_{soy\ milk})$	145	-0.65191	0.37565	-94.52747	-1.03066	0.05476
$\ln(p_{almond\ milk})$	145	-0.35368	0.13834	-51.28342	-0.71342	-0.03797
$\ln(p_{rice\ milk})$	145	-0.79428	0.19363	-115.17109	-1.12847	-0.31938
$\ln(p_{2\%\ milk})$	145	-1.23540	0.11578	-179.13237	-1.49410	-1.02623
$\ln(p_{1\%\ milk})$	145	-1.37851	0.11015	-199.88435	-1.59005	-1.20294
$\ln(p_{fat\text{-}free\ milk})$	145	-1.30290	0.11650	-188.91981	-1.54640	-1.06645
$\ln(p_{whole\ milk})$	145	-1.18322	0.11582	-171.56706	-1.43891	-0.98344

Note: $\ln(p_{soy\ milk})$=log of price of soy milk
$\ln(p_{almond\ milk})$=log of price of almond milk
$\ln(p_{rice\ milk})$=log of price of rice milk
$\ln(p_{2\%\ milk})$=log of price of 2% milk
$\ln(p_{1\%\ milk})$=log of price of 1% milk
$\ln(p_{fat\text{-}free\ milk})$=log of price of fat-free milk
$\ln(p_{whole\ milk})$=log of price of whole milk

Table 23 Correlation and Covariance Matrix of Log Prices of Seven Products

	$\ln(p_{soy\ milk})$	$\ln(p_{almond\ milk})$	$\ln(p_{rice\ milk})$	$\ln(p_{2\%\ milk})$	$\ln(p_{1\%\ milk})$	$\ln(p_{fat\text{-}free\ milk})$	$\ln(p_{whole\ milk})$
$\ln(p_{soy\ milk})$	1.00000	0.65145 <.0001	-0.20938 0.0115	-0.62267 <.0001	-0.59260 <.0001	-0.60123 <.0001	-0.63147 <.0001
$\ln(p_{almond\ milk})$	0.65145 <.0001	1.00000	0.00397 0.9622	-0.43102 <.0001	-0.38072 <.0001	-0.43315 <.0001	-0.42321 <.0001
$\ln(p_{rice\ milk})$	-0.20938 0.0115	0.00397 0.9622	1.00000	0.13445 0.1069	0.12123 0.1464	0.09626 0.2494	0.11933 0.1528
$\ln(p_{2\%\ milk})$	-0.62267 <.0001	-0.43102 <.0001	0.13445 0.1069	1.00000	0.97742 <.0001	0.98737 <.0001	0.98498 <.0001
$\ln(p_{1\%\ milk})$	-0.59260 <.0001	-0.38072 <.0001	0.12123 0.1464	0.97742 <.0001	1.00000	0.97329 <.0001	0.97566 <.0001
$\ln(p_{fat\text{-}free\ milk})$	-0.60123 <.0001	-0.43315 <.0001	0.09626 0.2494	0.98737 <.0001	0.97329 <.0001	1.00000	0.97668 <.0001
$\ln(p_{whole\ milk})$	-0.63147 <.0001	-0.42321 <.0001	0.11933 0.1528	0.98498 <.0001	0.97566 <.0001	0.97668 <.0001	1.00000

Note: "<.0001" represents P-values
$\ln(p_{soy\ milk})$=log of price of soy milk
$\ln(p_{almond\ milk})$=log of price of almond milk
$\ln(p_{rice\ milk})$=log of price of rice milk
$\ln(p_{2\%\ milk})$=log of price of 2% milk
$\ln(p_{1\%\ milk})$=log of price of 1% milk
$\ln(p_{fat\text{-}free\ milk})$=log of price of fat-free milk
$\ln(p_{whole\ milk})$=log of price of whole milk

CHAPTER IV

ESTIMATION AND RESULTS

The first step of hedonic metric approach is to estimate hedonic pricing models. Applying the model developed in chapter 2 and using the data file we constructed in chapter 3, we acquired estimates for all the variables considered for each product. In the following section, we discuss the hedonic regression results in detail.

Hedonic Pricing Model

The results of the hedonic regressions for almond milk, soy milk, and rice milk and four types of conventional milk are shown in Tables 24-29. The hedonic results generally conform to expectations. Both model forms fit well for dairy alternatives and regular fluid milk data. Regardless of the functional forms of hedonic regressions, almost all the nutritional variables are significant with only few exceptions. Compared with soy milk which has all nutritional variables significant in linear functional form, fat content is not significant for both almond milk and rice milk and calories is not significant for almond milk. For rice milk and soy milk, fat content has negative impact on monthly average prices which is understandable because people do not want to intake too many fats with the consideration of health and obesity issues. In linear hedonic regression, it seems that generally, Vitamin A and D has negatively affect people's willingness to pay. As shown in table 24, Vitamin D contributes negatively to prices of both soy and almond milk and Vitamin A has negative sign for rice milk. However, in log hedonic regression, those two nutritional variables are significant for all types of dairy alternatives. The possible reason for negative effects is that people do not think Vitamin A and D are necessary and important contents in dairy milk alternatives. The dummy variable Brands has negative sigh and significant effect as expected which means that prices of private label products are lower than national label products. Almost all package sizes dummy variables have positive and significant contribution to prices. To conclude, in general, people prefer smaller sizes at least not bigger than 64 oz. per package than 64 oz. per package. Coupon and deals have negative and significant effects on prices as expected. The effects of multi pack dummies on dairy alternatives have little inconsistency. The smaller units, the lower unit prices of products, which is consistent with soy milk as we can see that pack of 4 and 6 have positive influences meaning these two package units are preferred than pack of 1 for but for almond milk and rice consumers are inclined to by pack of 2 rather than other multi_packages. However, for almond milk and rice milk, the prices are higher for pack of 2 compared with pack of 1. Yearly dummies are significant for soy milk and rice milk especially from year 2004 to year 2009. However, the yearly dummies do not have significant effect on almond milk except for year 2008 and 2009. In terms of the log functional form, all the nutritional variables are significant for rice milk. Calories are not significant for all three types of dairy alternatives and fat content is not significant for soy milk. Similar to linear hedonic regressions, all package sizes dummy variables have positive and significant effects on prices. The effects of multi_pack dummies on prices in the log hedonic regression form are almost the same as in the linear hedonic form. Also, the yearly dummies are not showing much significance for almond milk. For regular milk products, we can witness that both functional forms fit the data very well. In term of 2% and 1% reduced fat milk, almost all the variables are significant at 0.1% level. One interesting result is that calcium has negative effect on prices of reduced fat milk but has positive effects on prices which is out of our expectations.

Table 24 Linear Hedonic Quality Attributes Estimates of DABs

soy milk				**almond milk**				**rice milk**			
Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value
Intercept	0.1648	0.0082	<.0001	Intercept	0.2370	0.0430	<.0001	Intercept	0.5850	0.0734	<.0001
x_{Kcal}	0.0003	0.0000	<.0001	x_{Kcal}	0.0003	0.0002	0.203	x_{Kcal}	-0.0001	0.0004	0.8808
x_{fat}	-0.0025	0.0010	0.016	x_{fat}	0.0111	0.0073	0.1316	x_{fat}	-0.1080	0.0109	<.0001
x_{VA}	0.0029	0.0002	<.0001	x_{VA}	0.0073	0.0007	<.0001	x_{VA}	-0.0140	0.0020	<.0001
x_{cal}	0.0008	0.0001	<.0001	x_{cal}	0.0040	0.0007	<.0001	x_{cal}	0.0043	0.0017	0.011
x_{VD}	-0.0007	0.0001	<.0001	x_{VD}	-0.0083	0.0009	<.0001	x_{VD}	0.0056	0.0007	<.0001
x_{fiber}	-0.0065	0.0014	<.0001	x_{fiber}	-0.0082	0.0035	0.0209	x_{fiber}	0.1504	0.0067	<.0001
x_{protein}	0.0248	0.0009	<.0001	x_{protein}	0.2115	0.0074	<.0001	x_{protein}	0.0079	0.0050	0.1112
D_{brands}	-0.0329	0.0024	<.0001					D_{brands}	-0.1780	0.0219	<.0001
$D_{\text{pkge\_size1}}$	0.4837	0.0174	<.0001	$D_{\text{pkge\_size1}}$	0.5736	0.0291	<.0001	$D_{\text{pkge\_size1}}$	0.2637	0.0808	0.0011
$D_{\text{pkge\_size2}}$	-0.0127	0.0321	0.6928	$D_{\text{pkge\_size2}}$	0.4747	0.0781	<.0001	$D_{\text{pkge\_size2}}$	0.2635	0.0330	<.0001
$D_{\text{pkge\_size3}}$	0.7651	0.0097	<.0001	$D_{\text{pkge\_size4}}$	0.0644	0.0529	0.2235	$D_{\text{pkge\_size3}}$	0.1410	0.0310	<.0001
$D_{\text{pkge\_size4}}$	0.1925	0.0963	0.0456	$D_{\text{pkge\_size5}}$	0.6390	0.0510	<.0001	$D_{\text{pkge\_size4}}$	0.2418	0.0282	<.0001
$D_{\text{pkge\_size5}}$	0.7511	0.0094	<.0001	$D_{\text{pkge\_size6}}$	0.2877	0.0470	<.0001	$D_{\text{pkge\_size5}}$	0.0259	0.0227	0.2542
$D_{\text{pkge\_size6}}$	0.9296	0.0172	<.0001	$D_{\text{pkge\_size9}}$	-0.1234	0.0145	<.0001	$D_{\text{pkge\_size6}}$	0.1023	0.0302	0.0007
$D_{\text{pkge\_size7}}$	0.6143	0.0175	<.0001					$D_{\text{pkge\_size7}}$	0.0021	0.0244	0.9327
$D_{\text{pkge\_size8}}$	1.0393	0.0089	<.0001								
$D_{\text{pkge\_size9}}$	1.0874	0.0258	<.0001								
$D_{\text{pkge\_size10}}$	0.1648	0.0036	<.0001								
$D_{\text{pkge\_size11}}$	-0.0387	0.0184	0.0353								
$D_{\text{pkge\_size12}}$	0.0507	0.0365	0.1647								
$D_{\text{pkge\_size14}}$	-0.2098	0.0194	<.0001								
D_{coupon}	-0.0226	0.0045	<.0001	D_{coupon}	-0.0431	0.0147	0.0034	D_{coupon}	-0.0916	0.0196	<.0001
D_{multi2}	-0.0410	0.0069	<.0001	D_{multi2}	0.0350	0.0799	0.6616	D_{multi2}	0.0479	0.0264	0.0703
D_{multi3}	-0.0678	0.0058	<.0001	D_{multi4}	-0.1444	0.0433	0.0009				
D_{multi4}	0.3328	0.0582	<.0001	D_{multi6}	-0.1336	0.0389	0.0006				
D_{multi6}	0.0787	0.0269	0.0034	D_{multi7}	-0.0021	0.0545	0.9695				
				D_{multi8}	-0.1252	0.0596	0.0358				

Table 24 Continued

soy milk				**almond milk**				**rice milk**			
Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value
D_{deals}	-0.0047	0.0028	0.1008	D_{deals}	-0.0310	0.0096	0.1306	D_{deals}	-0.0127	0.0094	0.1756
D_{year2014}	-0.0033	0.0043	0.4352	D_{year2014}	-0.0203	0.0134	0.5778	D_{year2014}	-0.0187	0.0144	0.1953
D_{year2013}	-0.0138	0.0043	0.0011	D_{year2013}	-0.0083	0.0150	0.8400	D_{year2013}	-0.0286	0.0149	0.0551
D_{year2012}	-0.0192	0.0042	<.0001	D_{year2012}	-0.0029	0.0146	0.1315	D_{year2012}	-0.0022	0.0142	0.8773
D_{year2011}	-0.0214	0.0041	<.0001	D_{year2011}	-0.0225	0.0149	0.3874	D_{year2011}	-0.0100	0.0147	0.4964
D_{year2010}	-0.0157	0.0041	0.0001	D_{year2010}	0.0132	0.0152	0.0094	D_{year2010}	0.0033	0.0158	0.8337
D_{year2009}	0.0354	0.0097	0.0003	D_{year2009}	0.0411	0.0158	0.0012	D_{year2009}	-0.0395	0.0154	0.0107
D_{year2008}	0.2175	0.0165	<.0001	D_{year2008}	0.0532	0.0164	0.2284	D_{year2008}	-0.0473	0.0160	0.0032
D_{year2007}	-0.0323	0.0042	<.0001	D_{year2007}	0.0211	0.0175	0.6948	D_{year2007}	-0.0726	0.0162	<.0001
D_{year2006}	-0.0531	0.0044	<.0001	D_{year2006}	0.0074	0.0189	0.4432	D_{year2006}	-0.0602	0.0167	0.0003
D_{year2005}	-0.0978	0.0198	<.0001	D_{year2005}	0.0179	0.0233	0.0509	D_{year2005}	-0.0756	0.0165	<.0001
D_{year2004}	0.0495	0.0291	0.0888	D_{year2004}	-0.0431	0.0221	0.1306	D_{year2004}	-0.0885	0.0163	<.0001
DF: 38	F Value: 1443.87		Pr > F: <.0001	DF: 31	F Value: 204.56		Pr > F: <.0001	DF: 29	F Value: 236.16		Pr >F:<.0001
RMSE 0.13550	Adj R^2 0.7987			RMSE: 0.1290	Adj R^2: 0.8175			RMSE: 0.0846	Adj R^2: 0.8907		

Table 25 Linear Hedonic Quality Estimates of 2% and 1% Milk Products

2% milk				**1% milk**			
Variables	Estimate	Std Error	P-value	Variable	Estimate	Standard	P-value
Intercept	0.5493	0.0151	<.0001	Intercept	-0.1870	0.0161	<.0001
x_{Kcal}	-0.0008	0.0001	<.0001	x_{Kcal}	0.0016	0.0001	<.0001
x_{fat}	0.0050	0.0008	<.0001	x_{fat}	-0.0124	0.0009	<.0001
x_{protein}	0.0383	0.0011	<.0001	x_{protein}	0.0259	0.0017	<.0001
x_{VA}	-0.0137	0.0007	<.0001	x_{VA}	0.0188	0.0005	<.0001
x_{cal}	-0.0153	0.0005	<.0001	x_{cal}	-0.0071	0.0003	<.0001
x_{VD}	0.0024	0.0001	<.0001	x_{VD}	0.0028	0.0004	<.0001
D_{brands}	-0.0142	0.0008	<.0001	D_{brands}	0.4607	0.0130	<.0001
$D_{\text{pkge\_size1}}$	4.7131	0.1284	<.0001	$D_{\text{pkge\_size1}}$	0.5646	0.0139	<.0001
$D_{\text{pkge\_size2}}$	0.3929	0.0045	<.0001	$D_{\text{pkge\_size2}}$	0.6653	0.0295	<.0001
$D_{\text{pkge\_size3}}$	0.6927	0.0127	<.0001	$D_{\text{pkge\_size3}}$	0.5145	0.0111	<.0001
$D_{\text{pkge\_size4}}$	0.2448	0.0097	<.0001	$D_{\text{pkge\_size4}}$	0.5282	0.0186	<.0001
$D_{\text{pkge\_size5}}$	0.3801	0.0101	<.0001	$D_{\text{pkge\_size5}}$	0.3219	0.0069	<.0001
$D_{\text{pkge\_size6}}$	0.6159	0.0287	<.0001	$D_{\text{pkge\_size6}}$	0.1791	0.0013	<.0001
$D_{\text{pkge\_size7}}$	0.4985	0.0046	<.0001	$D_{\text{pkge\_size7}}$	0.0499	0.0081	<.0001
$D_{\text{pkge\_size8}}$	0.4586	0.0078	<.0001	$D_{\text{pkge\_size8}}$	0.0852	0.0007	<.0001
$D_{\text{pkge\_size9}}$	0.3681	0.0018	<.0001	$D_{\text{pkge\_size9}}$	0.0754	0.0360	<.0001
$D_{\text{pkge\_size10}}$	0.1835	0.0170	<.0001	$D_{\text{pkge\_size10}}$	0.2332	0.0052	0.0363
$D_{\text{pkge\_size11}}$	0.7159	0.0389	<.0001	$D_{\text{pkge\_size11}}$	0.0577	0.0064	<.0001
$D_{\text{pkge\_size12}}$	0.1770	0.0012	<.0001	$D_{\text{pkge\_size12}}$	-0.1327	0.0134	<.0001
$D_{\text{pkge\_size13}}$	0.1197	0.0149	<.0001	$D_{\text{pkge\_size14}}$	-0.1870	0.0161	<.0001
$D_{\text{pkge\_size14}}$	0.2673	0.0321	<.0001				
$D_{\text{pkge\_size15}}$	0.2638	0.0344	<.0001				
$D_{\text{pkge\_size16}}$	0.0949	0.0009	<.0001				
$D_{\text{pkge\_size17}}$	0.0766	0.0321	0.0171				
$D_{\text{pkge\_size18}}$	0.1492	0.0057	<.0001				
$D_{\text{pkge\_size19}}$	0.0373	0.0106	0.0004				
D_{coupon}	-0.0208	0.0025	<.0001	D_{coupon}	-0.0146	0.0019	<.0001

Table 25 Continued

2% milk Variables	Estimate	Std Error	P-value	**1% milk** Variable	Estimate	Standard	P-value
D_{multi2}	-0.0543	0.0040	<.0001	D_{multi2}	-0.0395	0.0040	<.0001
D_{multi3}	0.0503	0.0100	<.0001	D_{multi3}	0.1009	0.0046	<.0001
D_{multi6}	-0.0245	0.0126	0.0516	D_{multi6}	0.3165	0.0304	<.0001
D_{multi7}	-0.2638	0.1285	0.0401	D_{multi7}	0.2274	0.0207	<.0001
D_{deals}	-0.0093	0.0013	<.0001	D_{deals}	-0.0141	0.0010	<.0001
D_{year2014}	0.0113	0.0020	<.0001	D_{year2014}	0.0154	0.0020	<.0001
D_{year2013}	-0.0031	0.0020	0.1131	D_{year2013}	-0.0060	0.0020	<.0001
D_{year2012}	-0.0145	0.0019	<.0001	D_{year2012}	-0.0107	0.0020	0.0030
D_{year2011}	-0.0134	0.0019	<.0001	D_{year2011}	-0.0051	0.0017	<.0001
D_{year2010}	-0.0411	0.0019	<.0001	D_{year2010}	-0.0323	0.0019	0.0029
D_{year2009}	-0.0542	0.0019	<.0001	D_{year2009}	-0.0539	0.0078	<.0001
D_{year2008}	-0.0229	0.0019	<.0001	D_{year2008}	-0.0028	0.0017	<.0001
D_{year2007}	-0.0417	0.0019	<.0001	D_{year2007}	-0.0173	0.0019	0.1012
D_{year2006}	-0.0718	0.0020	<.0001	D_{year2006}	-0.0498	0.0017	<.0001
D_{year2005}	-0.0701	0.0020	<.0001	D_{year2005}	-0.0520	0.0017	<.0001
D_{year2004}	-0.0748	0.0020	<.0001	D_{year2004}	-0.0545	0.0017	<.0001
DF: 43 F Value: 2151.21 Pr > F: <.0001 RMSE: 0.1284 Adj R^2: 0.4402				DF: 36 F Value: 1444.51 Pr > F: <.0001 RMSE: 0.0712 Adj R^2: 0.5126			

Table 26 Linear Hedonic Quality Estimates of Whole Milk and Fat-free Milk Products

whole milk				**fat-free milk**			
Variable	Estimate	Standard	P-value	Variable	Estimate	Standard	P-value
Intercept	-0.3287	0.0223	<.0001	Intercept	0.1298	0.0102	<.0001
x_{Kcal}	0.0052	0.0001	<.0001	x_{Kcal}	0.0004	0.0001	<.0001
x_{fat}	-0.1133	0.0019	<.0001	x_{fat}	-0.0016	0.0007	0.0168
x_{protein}	0.0140	0.0006	<.0001	x_{protein}	0.0051	0.0003	<.0001
x_{VA}	0.0042	0.0002	<.0001	x_{VA}	0.0012	0.0001	<.0001
x_{cal}	0.0016	0.0005	0.0004	x_{cal}	-0.0005	0.0001	<.0001
x_{VD}	0.0535	0.0007	<.0001	x_{VD}	-0.0013	0.0010	0.2106
D_{brands}	-0.0163	0.0008	<.0001	D_{brands}	-0.0097	0.0008	<.0001
$D_{\mathrm{pkge\_size1}}$	0.2580	0.0593	<.0001	$D_{\mathrm{pkge\_size1}}$	0.7955	0.0058	<.0001
$D_{\mathrm{pkge\_size2}}$	0.5022	0.0054	<.0001	$D_{\mathrm{pkge\_size2}}$	0.2563	0.0118	<.0001
$D_{\mathrm{pkge\_size3}}$	0.1775	0.0100	<.0001	$D_{\mathrm{pkge\_size3}}$	0.2685	0.0269	<.0001
$D_{\mathrm{pkge\_size4}}$	0.2333	0.0105	<.0001	$D_{\mathrm{pkge\_size4}}$	0.2806	0.0554	<.0001
$D_{\mathrm{pkge\_size5}}$	0.4032	0.0049	<.0001	$D_{\mathrm{pkge\_size5}}$	0.5783	0.0054	<.0001
$D_{\mathrm{pkge\_size6}}$	0.3508	0.0056	<.0001	$D_{\mathrm{pkge\_size6}}$	0.5051	0.0164	<.0001
$D_{\mathrm{pkge\_size7}}$	0.2344	0.0019	<.0001	$D_{\mathrm{pkge\_size7}}$	0.3474	0.0031	<.0001
$D_{\mathrm{pkge\_size8}}$	0.2397	0.0112	<.0001	$D_{\mathrm{pkge\_size8}}$	0.1781	0.0012	<.0001
$D_{\mathrm{pkge\_size9}}$	0.1671	0.0329	<.0001	$D_{\mathrm{pkge\_size9}}$	0.0819	0.0269	0.0023
$D_{\mathrm{pkge\_size10}}$	0.1141	0.0014	<.0001	$D_{\mathrm{pkge\_size10}}$	0.3477	0.0227	<.0001
$D_{\mathrm{pkge\_size11}}$	0.2144	0.0155	<.0001	$D_{\mathrm{pkge\_size11}}$	0.3097	0.0296	<.0001
$D_{\mathrm{pkge\_size12}}$	0.1412	0.0343	<.0001	$D_{\mathrm{pkge\_size12}}$	-0.1827	0.0459	<.0001
$D_{\mathrm{pkge\_size13}}$	0.1031	0.0124	<.0001	$D_{\mathrm{pkge\_size13}}$	0.0953	0.0008	<.0001
$D_{\mathrm{pkge\_size14}}$	0.0480	0.0010	<.0001	$D_{\mathrm{pkge\_size14}}$	0.0688	0.0278	0.0133
$D_{\mathrm{pkge\_size15}}$	0.0619	0.0057	<.0001	$D_{\mathrm{pkge\_size15}}$	0.1763	0.0047	<.0001
				$D_{\mathrm{pkge\_size16}}$	0.0508	0.0099	<.0001
D_{coupon}	-0.0204	0.0027	<.0001	D_{coupon}	-0.0148	0.0023	<.0001
D_{multi2}	-0.0487	0.0061	<.0001	D_{multi2}	-0.0581	0.0041	<.0001
D_{multi3}	0.0756	0.0141	<.0001				
D_{multi5}	-0.1771	0.0241	<.0001				

Table 26 Continued

whole milk				fat-free milk			
Variable	Estimate	Standard	P-value	Variable	Estimate	Standard	P-value
D_{multi6}	-0.5040	0.0686	<.0001				
D_{multi7}	-0.1304	0.0451	0.0039				
D_{deals}	-0.0092	0.0013	<.0001	D_{deals}	-0.0121	0.0012	<.0001
D_{year2014}	0.0091	0.0019	<.0001	D_{year2014}	0.0185	0.0018	<.0001
D_{year2013}	-0.0070	0.0019	0.0003	D_{year2013}	0.0106	0.0017	<.0001
D_{year2012}	-0.0101	0.0019	<.0001	D_{year2012}	-0.0032	0.0018	0.0722
D_{year2011}	-0.0142	0.0017	<.0001	D_{year2011}	-0.0057	0.0017	0.0008
D_{year2010}	-0.0408	0.0019	<.0001	D_{year2010}	-0.0284	0.0018	<.0001
D_{year2009}	-0.0578	0.0020	<.0001	D_{year2009}	-0.0423	0.0026	<.0001
D_{year2008}	-0.0243	0.0018	<.0001	D_{year2008}	-0.0081	0.0017	<.0001
D_{year2007}	-0.0340	0.0034	<.0001	D_{year2007}	-0.0248	0.0020	<.0001
D_{year2006}	-0.0703	0.0020	<.0001	D_{year2006}	-0.0603	0.0017	<.0001
D_{year2005}	-0.0660	0.0020	<.0001	D_{year2005}	-0.0609	0.0018	<.0001
D_{year2004}	-0.0709	0.0020	<.0001	D_{year2004}	-0.06428	0.00174	<.0001
DF: 43 F Value: 3116.00 Pr > F: <.0001 RMSE: 0.1185 Adj R^2: 0.5538				DF: 37 F Value: 1906.44 Pr > F: <.0001 RMSE: 0.1108 Adj R^2: 0.4441			

Table 27 Log Hedonic Quality Attributes Estimates of DABs

soymilk				**almond milk**				**rice milk**			
Variables	Estimat	Std Error	P-value	Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value
Intercept	-1.3413	0.0165	<.0001	Intercept	-0.9270	0.0635	<.0001	Intercept	-0.6950	0.1463	<.0001
x_{Kcal}	0.0001	0.0001	0.0907	x_{Kcal}	-0.0002	0.0003	0.4925	x_{Kcal}	-0.0018	0.0007	0.0156
x_{fat}	0.0033	0.0021	0.1113	x_{fat}	0.0649	0.0108	<.0001	x_{fat}	-0.1857	0.0217	<.0001
x_{VA}	0.0056	0.0004	<.0001	x_{VA}	0.0060	0.0010	<.0001	x_{VA}	-0.0248	0.0040	<.0001
x_{cal}	0.0025	0.0003	<.0001	x_{cal}	0.0056	0.0011	<.0001	x_{cal}	0.0098	0.0033	0.0035
x_{VD}	-0.0011	0.0003	<.0001	x_{VD}	-0.0136	0.0013	<.0001	x_{VD}	0.0141	0.0013	<.0001
x_{fiber}	-0.0187	0.0029	<.0001	x_{fiber}	-0.0089	0.0052	0.0888	x_{fiber}	0.1401	0.0133	<.0001
x_{protein}	0.0357	0.0018	<.0001	x_{protein}	0.1514	0.0109	<.0001	x_{protein}	0.0136	0.0099	0.17
D_{brands}	-0.0623	0.0048	<.0001	$D_{\text{pkge\_size1}}$	0.7588	0.0430	<.0001	D_{brands}	-0.3464	0.0436	<.0001
$D_{\text{pkge\_size1}}$	0.8931	0.0349	<.0001	$D_{\text{pkge\_size2}}$	0.6859	0.1154	<.0001	$D_{\text{pkge\_size1}}$	0.7094	0.1610	<.0001
$D_{\text{pkge\_size2}}$	0.1949	0.0643	0.0024	$D_{\text{pkge\_size4}}$	0.0890	0.0781	0.2543	$D_{\text{pkge\_size2}}$	0.6611	0.0657	<.0001
$D_{\text{pkge\_size3}}$	1.1565	0.0194	<.0001	$D_{\text{pkge\_size5}}$	0.8575	0.0753	<.0001	$D_{\text{pkge\_size3}}$	0.5649	0.0618	<.0001
$D_{\text{pkge\_size4}}$	0.4572	0.1926	0.0176	$D_{\text{pkge\_size6}}$	0.4334	0.0695	<.0001	$D_{\text{pkge\_size4}}$	0.7195	0.0563	<.0001
$D_{\text{pkge\_size5}}$	1.1361	0.0188	<.0001	$D_{\text{pkge\_size9}}$	-0.1833	0.0214	<.0001	$D_{\text{pkge\_size5}}$	0.1699	0.0452	0.0002
$D_{\text{pkge\_size6}}$	1.1577	0.0345	<.0001					$D_{\text{pkge\_size6}}$	0.4754	0.0601	<.0001
$D_{\text{pkge\_size7}}$	0.8083	0.0350	<.0001					$D_{\text{pkge\_size7}}$	0.1473	0.0485	0.0025
$D_{\text{pkge\_size8}}$	1.3573	0.0178	<.0001								
$D_{\text{pkge\_size9}}$	1.3775	0.0515	<.0001								
$D_{\text{pkge\_size10}}$	0.3612	0.0071	<.0001								
$D_{\text{pkge\_size11}}$	-0.0823	0.0368	0.0254								
$D_{\text{pkge\_size12}}$	-0.0446	0.0729	0.5407								
$D_{\text{pkge\_size14}}$	-0.7241	0.0389	<.0001								
D_{coupon}	-0.0423	0.0091	<.0001	D_{coupon}	-0.0316	0.0217	0.1456	D_{coupon}	-0.2683	0.0391	<.0001
D_{multi2}	-0.0603	0.0137	<.0001	D_{multi2}	0.1594	0.1181	0.1772	D_{multi2}	0.0651	0.0526	0.2163
D_{multi3}	-0.1840	0.0115	<.0001	D_{multi4}	-0.1513	0.0640	0.0182				
D_{multi4}	0.2898	0.1164	0.0128	D_{multi6}	-0.2495	0.0575	<.0001				
D_{multi6}	-0.0393	0.0538	0.4652	D_{multi7}	-0.0121	0.0805	0.8805				

Table 27 Continued

soy milk Variable	Estimate	Std Error	P-value
D_{deals}	-0.0134	0.0057	0.0185
D_{year2014}	-0.0076	0.0086	0.3757
D_{year2013}	-0.0325	0.0085	0.0001
D_{year2012}	-0.0390	0.0083	<.0001
D_{year2011}	-0.0480	0.0082	<.0001
D_{year2010}	-0.0470	0.0082	<.0001
D_{year2009}	0.0539	0.0194	0.0054
D_{year2008}	0.3445	0.0331	<.0001
D_{year2007}	-0.0766	0.0084	<.0001
D_{year2006}	-0.1117	0.0089	<.0001
D_{year2005}	-0.0816	0.0396	0.0396
D_{year2004}	-0.0107	0.0582	0.8546

DF: 38 F Value: 749.46 Pr > F: <.0001
RMSE: 0.1923 Adj R^2: 0.7400

almond milk Variables	Estimate	Std Error	P-value
D_{multi8}	-0.1600	0.0881	0.0694
D_{deals}	-0.0255	0.0142	0.0728
D_{year2014}	-0.0040	0.0198	0.8400
D_{year2013}	0.0166	0.0221	0.4523
D_{year2012}	0.0200	0.0215	0.3538
D_{year2011}	-0.0066	0.0221	0.7647
D_{year2010}	0.0546	0.0225	0.0154
D_{year2009}	0.0766	0.0234	0.0011
D_{year2008}	0.1107	0.0242	<.0001
D_{year2007}	0.0724	0.0258	0.0051
D_{year2006}	0.0443	0.0279	0.1129
D_{year2005}	0.0915	0.0344	0.0079
D_{year2004}	-0.0128	0.0326	0.6940

DF: 31 F Value: 112.48 Pr > F: <.0001
RMSE: 0.2002 Adj R^2: 0.6849

rice milk Variables	Estimate	Std Error	P-value
D_{deals}	-0.0347	0.0187	0.0643
D_{year2014}	-0.0423	0.0287	0.1409
D_{year2013}	-0.0631	0.0297	0.0338
D_{year2012}	-0.0396	0.0283	0.1627
D_{year2011}	-0.0423	0.0293	0.1494
D_{year2010}	0.0106	0.0315	0.7368
D_{year2009}	-0.0831	0.0308	0.007
D_{year2008}	-0.0942	0.0319	0.0032
D_{year2007}	-0.1244	0.0322	0.0001
D_{year2006}	-0.1015	0.0333	0.0024
D_{year2005}	-0.1347	0.0329	<.0001
D_{year2004}	-0.1671	0.0325	<.0001

DF: 29 F Value: 151.88 Pr > F: <.0001
RMSE: 0.1685 Adj R^2: 0.8394

Table 28 Log Hedonic Quality Estimates of 2% and 1% Milk Products

2% milk				**1%milk**			
Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value
Intercept	-0.8102	0.0306	<.0001	Intercept	-2.7819	0.0568	<.0001
x_{Kcal}	-0.0022	0.0001	<.0001	x_{Kcal}	0.0053	0.0004	<.0001
x_{fat}	0.0060	0.0017	0.0003	x_{fat}	-0.0478	0.0033	<.0001
x_{protein}	0.1165	0.0023	<.0001	x_{protein}	0.0799	0.0060	<.0001
x_{VA}	-0.0264	0.0015	<.0001	x_{VA}	0.0589	0.0017	<.0001
x_{cal}	-0.0428	0.0010	<.0001	x_{cal}	-0.0210	0.0011	<.0001
x_{VD}	0.0074	0.0002	<.0001	x_{VD}	0.0046	0.0014	<.0001
D_{brands}	-0.0406	0.0017	<.0001	D_{brands}	1.1251	0.0457	<.0001
$D_{\text{pkge\_size1}}$	3.2367	0.2610	<.0001	$D_{\text{pkge\_size1}}$	1.3765	0.0490	<.0001
$D_{\text{pkge\_size2}}$	0.9903	0.0091	<.0001	$D_{\text{pkge\_size2}}$	1.1757	0.1043	<.0001
$D_{\text{pkge\_size3}}$	1.5163	0.0258	<.0001	$D_{\text{pkge\_size3}}$	1.3205	0.0393	<.0001
$D_{\text{pkge\_size4}}$	0.7680	0.0196	<.0001	$D_{\text{pkge\_size4}}$	1.1903	0.0657	<.0001
$D_{\text{pkge\_size5}}$	1.0305	0.0206	<.0001	$D_{\text{pkge\_size5}}$	0.9907	0.0242	<.0001
$D_{\text{pkge\_size6}}$	1.4012	0.0584	<.0001	$D_{\text{pkge\_size6}}$	0.6274	0.0044	<.0001
$D_{\text{pkge\_size7}}$	1.2354	0.0095	<.0001	$D_{\text{pkge\_size7}}$	0.2850	0.0287	<.0001
$D_{\text{pkge\_size8}}$	1.1369	0.0158	<.0001	$D_{\text{pkge\_size8}}$	0.3515	0.0025	<.0001
$D_{\text{pkge\_size9}}$	1.0067	0.0037	<.0001	$D_{\text{pkge\_size9}}$	0.3448	0.1272	<.0001
$D_{\text{pkge\_size10}}$	0.6408	0.0346	<.0001	$D_{\text{pkge\_size10}}$	0.8206	0.0182	0.0067
$D_{\text{pkge\_size11}}$	1.4743	0.0789	<.0001	$D_{\text{pkge\_size11}}$	0.2792	0.0225	<.0001
$D_{\text{pkge\_size12}}$	0.6081	0.0024	<.0001	$D_{\text{pkge\_size12}}$	-1.1616	0.0473	<.0001
$D_{\text{pkge\_size13}}$	0.2989	0.0302	<.0001	$D_{\text{pkge\_size14}}$	-2.7819	0.0568	<.0001
$D_{\text{pkge\_size14}}$	0.8056	0.0653	<.0001				
$D_{\text{pkge\_size15}}$	0.8351	0.0698	<.0001				
$D_{\text{pkge\_size16}}$	0.3689	0.0017	<.0001				
$D_{\text{pkge\_size17}}$	0.3371	0.0653	<.0001				
$D_{\text{pkge\_size18}}$	0.5269	0.0115	<.0001				
$D_{\text{pkge\_size19}}$	0.2012	0.0216	<.0001				
D_{coupon}	-0.0964	0.0052	<.0001	D_{coupon}	-0.2333	0.0142	<.0001

Table 28 Continued

2% milk				**1%milk**			
Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value
D_{multi2}	-0.3122	0.0081	<.0001	D_{multi2}	0.3389	0.0161	<.0001
D_{multi3}	0.1550	0.0204	<.0001	D_{multi3}	0.4186	0.1072	<.0001
D_{multi6}	0.0700	0.0255	0.0062	D_{multi6}	0.4233	0.0732	<.0001
D_{multi7}	-0.4437	0.2611	0.0893	D_{multi7}	-0.0833	0.0034	<.0001
D_{deals}	-0.0650	0.0026	<.0001	D_{deals}	0.0894	0.0072	<.0001
D_{year2014}	0.0566	0.0040	<.0001	D_{year2014}	0.0027	0.0072	0.7040
D_{year2013}	0.0098	0.0040	0.0136	D_{year2013}	-0.0146	0.0071	0.0400
D_{year2012}	-0.0273	0.0039	<.0001	D_{year2012}	0.0060	0.0061	0.3232
D_{year2011}	-0.0319	0.0039	<.0001	D_{year2011}	-0.1124	0.0068	<.0001
D_{year2010}	-0.1253	0.0039	<.0001	D_{year2010}	-0.2117	0.0274	<.0001
D_{year2009}	-0.1783	0.0039	<.0001	D_{year2009}	0.0335	0.0060	<.0001
D_{year2008}	-0.0367	0.0038	<.0001	D_{year2008}	-0.0412	0.0066	<.0001
D_{year2007}	-0.1104	0.0038	<.0001	D_{year2007}	-0.1872	0.0058	<.0001
D_{year2006}	-0.2384	0.0040	<.0001	D_{year2006}	-0.1680	0.0061	<.0001
D_{year2005}	-0.2233	0.0040	<.0001	D_{year2005}	-0.1787	0.0061	<.0001
D_{year2004}	-0.2427	0.0040	<.0001	D_{year2004}	-0.2333	0.0142	<.0001
DF: 43 F Value: 4543.94 Pr > F:<.0001				DF: 36 F Value: 1531.86 Pr > F: <.0001			
RMSE: 0.26094 Adj R^2: 0.6243				RMSE: 0.25427 Adj R^2: 0.5272			

Table 29 Log Hedonic Quality Estimates of Whole Milk and Fat-free Milk Products

whole milk				**fat-free milk**			
Variable	Parameter	Standard	P-value	Variable	Parameter	Standard	P-value
Intercept	-3.4331	0.0445	<.0001	Intercept	-1.6814	0.0257	<.0001
x_{Kcal}	0.0098	0.0003	<.0001	x_{Kcal}	0.0025	0.0002	<.0001
x_{fat}	-0.1839	0.0039	<.0001	x_{fat}	-0.0061	0.0017	0.0003
x_{VA}	0.0254	0.0013	<.0001	x_{VA}	0.0043	0.0008	<.0001
x_{cal}	0.0220	0.0005	<.0001	x_{cal}	0.0050	0.0004	<.0001
x_{VD}	-0.0045	0.0009	<.0001	x_{VD}	-0.0005	0.0003	0.1498
x_{protein}	0.1480	0.0013	<.0001	x_{protein}	-0.0333	0.0026	<.0001
D_{brands}	-0.0384	0.0016	<.0001	D_{brands}	-0.0264	0.0021	<.0001
$D_{\text{pkge\_size1}}$	0.6397	0.1182	<.0001	$D_{\text{pkge\_size1}}$	1.3914	0.0147	<.0001
$D_{\text{pkge\_size2}}$	0.8763	0.0107	<.0001	$D_{\text{pkge\_size2}}$	0.8139	0.0297	<.0001
$D_{\text{pkge\_size3}}$	0.5044	0.0200	<.0001	$D_{\text{pkge\_size3}}$	0.8712	0.0680	<.0001
$D_{\text{pkge\_size4}}$	0.5959	0.0210	<.0001	$D_{\text{pkge\_size4}}$	1.0430	0.1401	<.0001
$D_{\text{pkge\_size5}}$	0.8492	0.0098	<.0001	$D_{\text{pkge\_size5}}$	1.3589	0.0136	<.0001
$D_{\text{pkge\_size6}}$	0.7191	0.0112	<.0001	$D_{\text{pkge\_size6}}$	1.2257	0.0414	<.0001
$D_{\text{pkge\_size7}}$	0.5987	0.0037	<.0001	$D_{\text{pkge\_size7}}$	1.0189	0.0079	<.0001
$D_{\text{pkge\_size8}}$	0.6012	0.0224	<.0001	$D_{\text{pkge\_size8}}$	0.6308	0.0031	<.0001
$D_{\text{pkge\_size9}}$	0.5192	0.0656	<.0001	$D_{\text{pkge\_size9}}$	0.2804	0.0680	<.0001
$D_{\text{pkge\_size10}}$	0.3694	0.0027	<.0001	$D_{\text{pkge\_size10}}$	0.9848	0.0573	<.0001
$D_{\text{pkge\_size11}}$	0.4528	0.0308	<.0001	$D_{\text{pkge\_size11}}$	0.9327	0.0749	<.0001
$D_{\text{pkge\_size12}}$	0.4081	0.0683	<.0001	$D_{\text{pkge\_size12}}$	-0.3710	0.1160	0.0014
$D_{\text{pkge\_size13}}$	0.3606	0.0247	<.0001	$D_{\text{pkge\_size13}}$	0.3803	0.0021	<.0001
$D_{\text{pkge\_size14}}$	0.2053	0.0020	<.0001	$D_{\text{pkge\_size14}}$	0.3430	0.0702	<.0001
$D_{\text{pkge\_size15}}$	0.2470	0.0113	<.0001	$D_{\text{pkge\_size15}}$	0.5997	0.0120	<.0001
				$D_{\text{pkge\_size16}}$	0.2843	0.0250	<.0001
D_{coupon}	-0.1081	0.0055	<.0001	D_{coupon}	-0.1043	0.0058	<.0001
D_{multi2}	-0.2169	0.0122	<.0001	D_{multi2}	-0.2792	0.0103	<.0001
D_{multi3}	0.1976	0.0281	<.0001				
D_{multi5}	-0.3895	0.0480	<.0001				

Table 29 Continued

2% milk				**1%milk**			
Variables	Estimate	Std Error	P-value	Variables	Estimate	Std Error	P-value
D_{multi6}	-0.7621	0.1369	<.0001				
D_{multi7}	-0.1700	0.0900	0.0590				
D_{deals}	-0.0577	0.0026	<.0001	D_{deals}	-0.0687	0.0030	<.0001
D_{year2014}	0.0509	0.0038	<.0001	D_{year2014}	0.0807	0.0045	<.0001
D_{year2013}	-0.0089	0.0038	0.0209	D_{year2013}	0.0463	0.0044	<.0001
D_{year2012}	-0.0196	0.0038	<.0001	D_{year2012}	-0.0041	0.0045	0.3593
D_{year2011}	-0.0201	0.0034	<.0001	D_{year2011}	-0.0142	0.0043	0.0009
D_{year2010}	-0.1134	0.0038	<.0001	D_{year2010}	-0.1055	0.0046	<.0001
D_{year2009}	-0.1768	0.0039	<.0001	D_{year2009}	-0.1523	0.0066	<.0001
D_{year2008}	-0.0273	0.0036	<.0001	D_{year2008}	-0.0020	0.0042	0.6373
D_{year2007}	-0.1053	0.0068	<.0001	D_{year2007}	-0.0744	0.0050	<.0001
D_{year2006}	-0.2053	0.0039	<.0001	D_{year2006}	-0.2251	0.0044	<.0001
D_{year2005}	-0.1815	0.0039	<.0001	D_{year2005}	-0.2100	0.0045	<.0001
D_{year2004}	-0.2000	0.0039	<.0001	D_{year2004}	-0.2316	0.0044	<.0001
DF: 43 F Value: 4543.94 Pr > F: <.0001 RMSE: 0.2609 Adj R^2: 0.6243				DF: 37 F Value: 2773.87 Pr >F: <.0001 RMSE: 0.2801 Adj R^2: 0.5376			

Barten Synthetic Demand Model

After acquiring the coefficient estimates from linear and log hedonic pricing models, we calculate the value added terms and pair-wised difference in-between to obtain the hedonic distance matrix. Thus far, we gathered all the parameters ready for estimating the Barten synthetic demand model. Expenditure, own-price and cross-price elasticity (both uncompensated and compensated) are estimated for the seven products over the 144-month period. Accordingly, we drop one equation for estimation purposes (Barten, 1969). Barten (1969) shows that parameter estimates are invariant to the equation dropped and the dropped parameters can be calculated from the adding-up restrictions of the model. Parameter estimates for each of aforementioned models are reported.

We examined presence of possible autocorrelation (serial correlation) through the autocorrelation and partial autocorrelation function generated for each series. It must be emphasized that the popular Durbin-Watson statistic could not be used to test for autocorrelation due the presence of lag of dependent variable (expenditure share and quantity in our work) in calculating the Divisia quantity index and average of budget shares in our Barten model. Alternatively, the test statistic suggested for such situations, i.e. Durbin-h statistic could not be used due to the fact that Durbin-h statistic broke down for situations where the product of the number of observations and variance of the estimated coefficient exceeded unity (Dharmasena, 2010).

Calculated autocorrelation and partial autocorrelation functions of the residuals of all dairy milk alternative beverages indicated the presence of possible serial correlation (this was expected to be the case given the time-series nature of the data set). A close study of the data indicated the presence of third-order or fourth-order autoregressive process of disturbance terms in the system. Therefore, each system was fitted with first- second-, third- and fourth- order autoregressive process of disturbance terms and significance of autocorrelation coefficient was investigated. Through such exercise, we found that disturbance terms behave as an *AR(5)* process. Thus Barten synthetic model was fitted assuming the disturbance process was:

(48) $e_{it} = \rho_{i1}e_{i,t-1} + \rho_{i2}e_{i,t-2} + \rho_{i3}e_{i,t-3} + \rho_{i4}e_{i,t-4} + \rho_{i4}e_{i,t-5} + u_{it}$

where ρ_{i1}, ρ_{i2}, ρ_{i3} , $\rho_{i4}$and ρ_{i5}represents first, second, third and fourth order autoregressive parameters respectively. The white-noise disturbance term is denoted by u_{it} which is independently and identically distributed with zero mean and constant variance. Finally, the estimating form of the re-parameterized Barten synthetic model taking into account *AR(5)* disturbances can be written as follows:

(49)

$w_{it}d\ln q_{it} = \rho_1(w_{it}d\ln q_{it})_{t-1} + \rho_2(w_{it}d\ln q_{it})_{t-2} + \rho_3(w_{it}d\ln q_{it})_{t-3} + \rho_4(w_{it}d\ln q_{it})_{t-4} +$

$\rho_5(w_{it}d\ln q_{it})_{t-5} + (\beta_i + \lambda w_{it})d\ln Q_t + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_i(\delta_{ii} - w_{it})\right)d\ln p_{it} +$

$\sum_{i\neq j}\left[\mathrm{c}_h d_{ij}^h + \mathrm{c}_{nn} d_{ij}^{nn} - \mu w_{it}\left(\delta_{ij} - w_{jt}\right)\right]d\ln p_{jt} -$

$$\rho_1\{(\beta_i + \lambda w_{it-1})d\ln Q_{t-1} + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_{it-1}(\delta_{ii} - w_{it-1})\right)d\ln p_{it-1}$$

$$+ \sum_{i\neq j}\left[\mathrm{c}_h d_{ij}^h + \mathrm{c}_{nn} d_{ij}^{nn} - \mu w_{it-1}\left(\delta_{ij} - w_{jt-1}\right)\right]d\ln p_{jt-1}\} -$$

$$\rho_2\{(\beta_i + \lambda w_{it-2})d\ln Q_{t-2} + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_{it-2}(\delta_{ii} - w_{it-2})\right)d\ln p_{it-2}$$

$$+ \sum_{i\neq j}\left[\mathrm{c}_h d_{ij}^h + \mathrm{c}_{nn} d_{ij}^{nn} - \mu w_{it-2}\left(\delta_{ij} - w_{jt-2}\right)\right]d\ln p_{jt-2}\} -$$

$$\rho_3\{(\beta_i + \lambda w_{it-3})d\ln Q_{t-3} + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_{it-3}(\delta_{ii} - w_{it-3})\right)d\ln p_{it-3}$$

$$+ \sum_{i\neq j}\left[\mathrm{c}_h d_{ij}^h + \mathrm{c}_{nn} d_{ij}^{nn} - \mu w_{it-3}\left(\delta_{ij} - w_{jt-3}\right)\right]d\ln p_{jt-3}\} -$$

$$\rho_4\{(\beta_i + \lambda w_{it-4})d\ln Q_{t-4} + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_{it-4}(\delta_{ii} - w_{it-4})\right)d\ln p_{it-4}$$

$$+ \sum_{i\neq j}\left[c_h d_{ij}^h + c_{nn} d_{ij}^{nn} - \mu w_{it-4}\left(\delta_{ij} - w_{jt-4}\right)\right]d\ln p_{jt-4}\} -$$

$$\rho_5\{(\beta_i + \lambda w_{it-5})d\ln Q_{t-5} + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_{it-5}(\delta_{ii} - w_{it-5})\right)d\ln p_{it-5}$$

$$+ \sum_{i\neq j}\left[c_h d_{ij}^h + c_{nn} d_{ij}^{nn} - \mu w_{it-5}\left(\delta_{ij} - w_{jt-5}\right)\right]d\ln p_{jt-5}\} + e_{it}$$

Trends in Budget Shares and Seasonality

We employed the following version of Barten synthetic model with an additive disturbance term and a seasonal adjustment done using quarterly seasonal dummies:

(47) $w_{it}d\ln q_{it} = (\beta_i + \lambda w_{it})d\ln Q_t + \alpha_0 d\ln p_{it} + \left(\alpha_1 x_i^s + \alpha_2 x_i^c - \mu w_i(\delta_{ii} - w_{it})\right)d\ln p_{it} +$
$\sum_{i\neq j}\left[c_h d_{ij}^h + c_{nn} d_{ij}^{nn} - \mu w_{it}\left(\delta_{ij} - w_{jt}\right)\right]d\ln p_{jt} + \sum_{j=1}^{3} d_j Q_{ijt} + e_{it}$

where i = (1,2,...,7) indexes seven dairy milk alternative beverage and conventional milk categories in the system, t indexes the time in months, i.e. t = (1,2,3,..., 145) p_{jt} is monthly average prices for each milk product considered in study, q_{it} is quantity (oz.) consumed in each milk product, Q_{ijt} is the quarterly dummy used to capture the seasonality pertaining to four quarters of the year. Monthly budget share of each DMAB consumed is denoted by w_{it} where $w_{it} = \frac{p_{jt}q_{it}}{m}$. Additive disturbance term is denoted by e_{it}.

Figures 13 through 19 shows the trends in budget shares of milk products considered in our study from January 2004 through December 2015 (on a monthly basis). Budget shares pertaining to rice milk and soy milk trend slightly down over the period. Almond milk exhibits upward trends in budget shares. Soy milk show a upward trend from the beginning of year 2006 but then turn downward from year 2008 to 2010. Whole milk budget shares show a slight upward after year 2012. The budget share associated with fat-free milk does not trend on either direction from

January 2004 to the end of 2009; thereafter it shows a slight downward trend.

Visual observation reveals that 1% milk, 2% milk, fat-free milk and whole milk show seasonality in the movement of budget shares over the sample period. More specifically, consumption of 1% milk, whole milk and almond milk is comparatively high in the second and third quarters and low in fourth and first quarters, 2% milk is consumed heavily during the fourth and first quarters (associated more with winter and holiday season), and relatively low consumption is observed during second and third quarters.

Joint hypotheses test for seasonal dummies, lambda and mu are shown in table 29. Significance of seasonal dummies for almond milk, soy milk and fat-free milk confirms the presence of seasonality in the data set, which is somewhat different from the results of visual observations.

Examination of individual seasonal dummies associated with each milk product revealed the following. More almond milk and soy milk are consumed in quarters 1, 2, and 3 compared to the fourth quarter. The most is consumed in the first quarter. This result is in accordance with Figure 13 and 14. However, rice milk and 2% fat milk are consumed less in the 1,2 and 3 quarters compared to the fourth quarter but the effects are not significant. Again, this result reinforced the budget share trends graphed in Figure 15 and 17. 1% low-fat milk is consumed less in the 1,2 and 3 quarters too but are very significant. This result is further confirmed through the joint hypothesis test we performed for the quarterly dummies of 1% low-fat milk. Again, budget share trends shown in the figure 16 provide evidence to support this result.

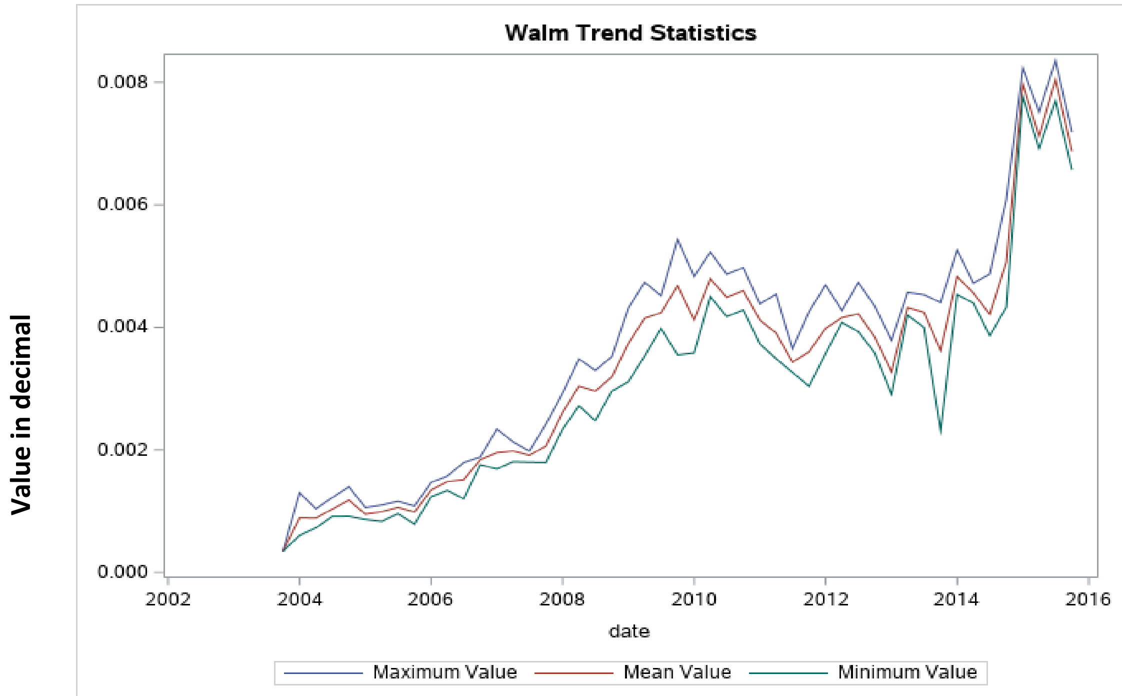


Figure 13 Trend Statistics in Budget Share: Almond Milk (Walm)

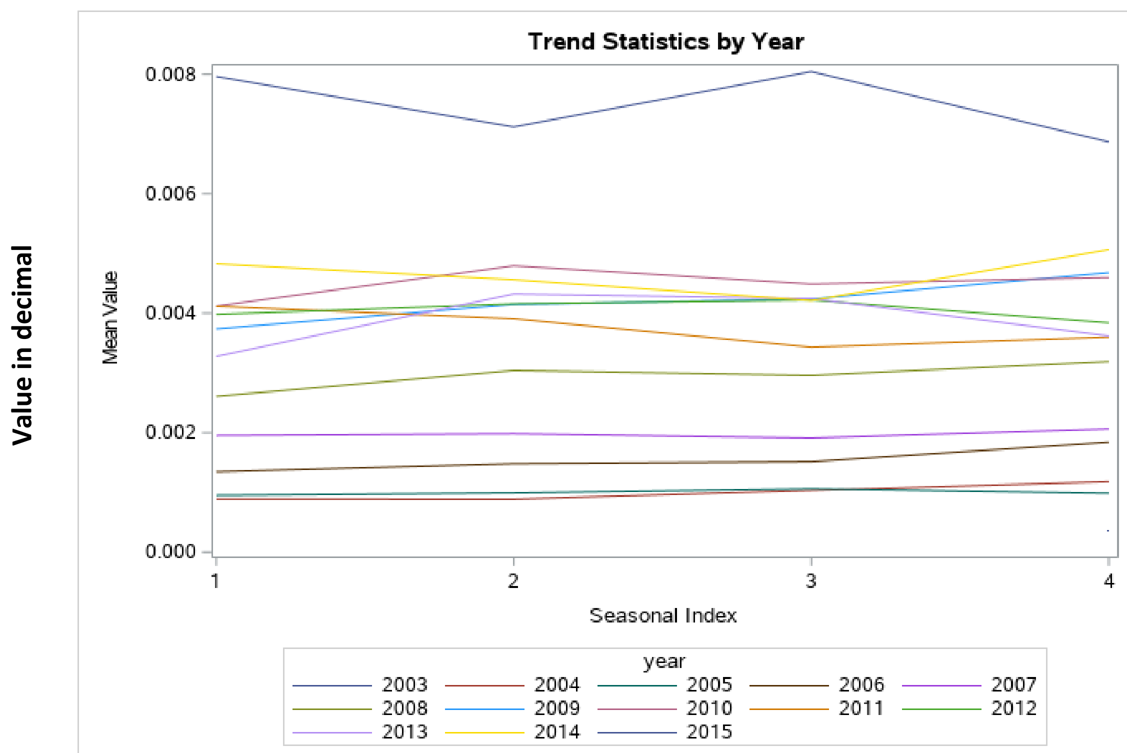


Figure 14 Trend Statistics by Year in Budget Share: Almond Milk (Walm)

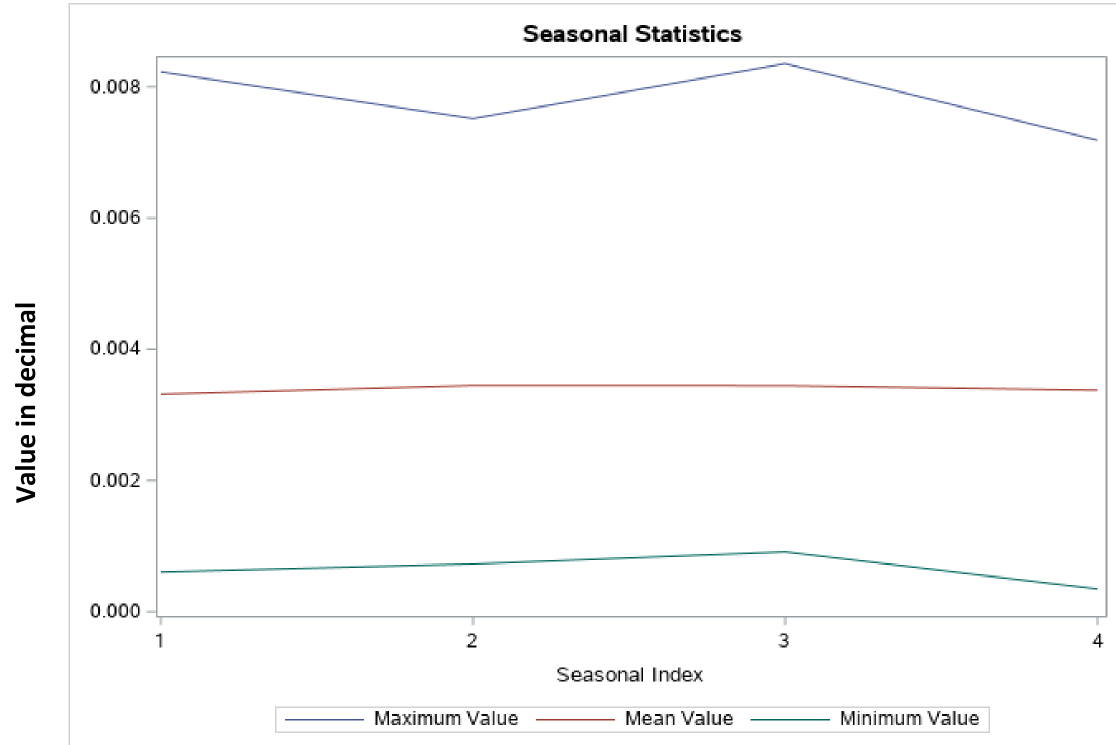


Figure 15 Seasonal Statistics in Budget Share: Almond Milk (Walm)

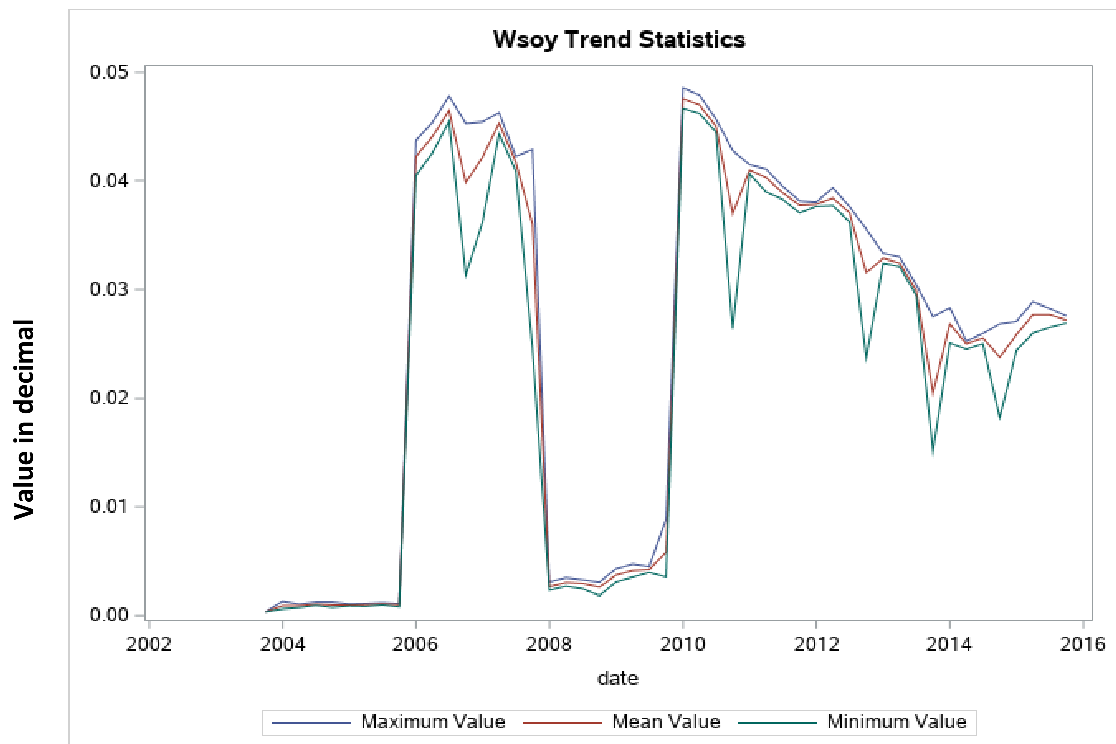


Figure 16 Trend Statistics in Budget Share: Soy Milk (Wsoy)

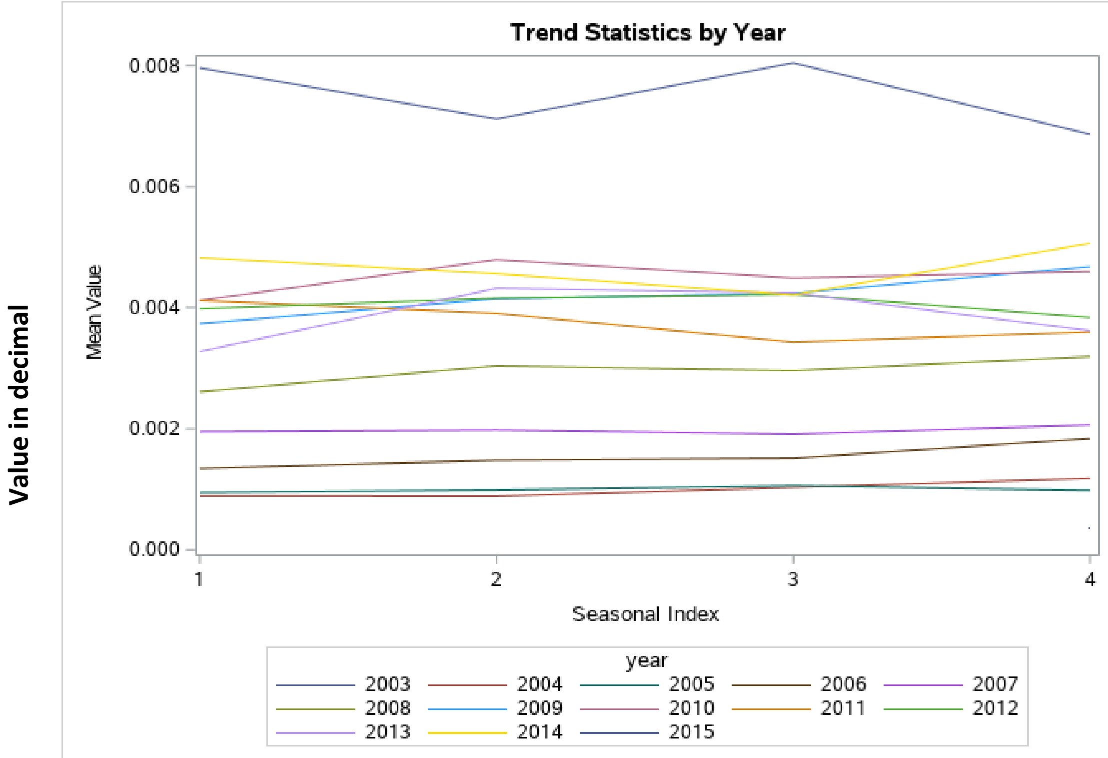


Figure 17 Trend Statistics by Year in Budget Share: Soy Milk (Wsoy)

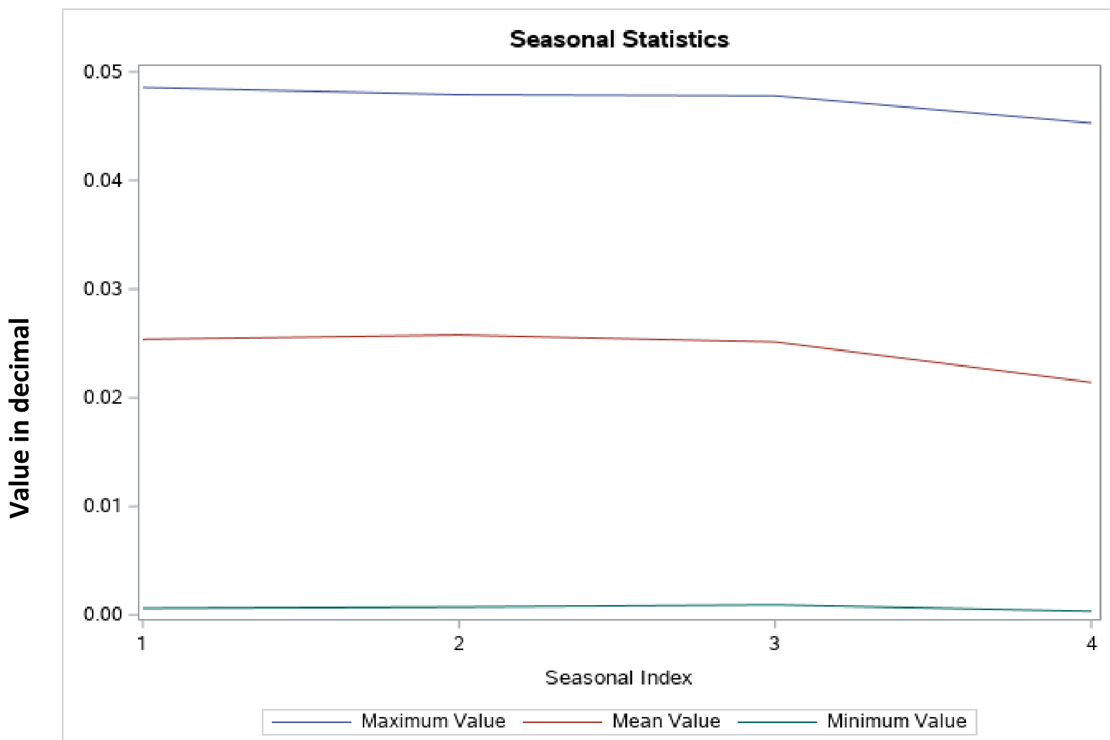


Figure 18 Seasonal Statistics in Budget Share: Soy Milk (Wsoy)

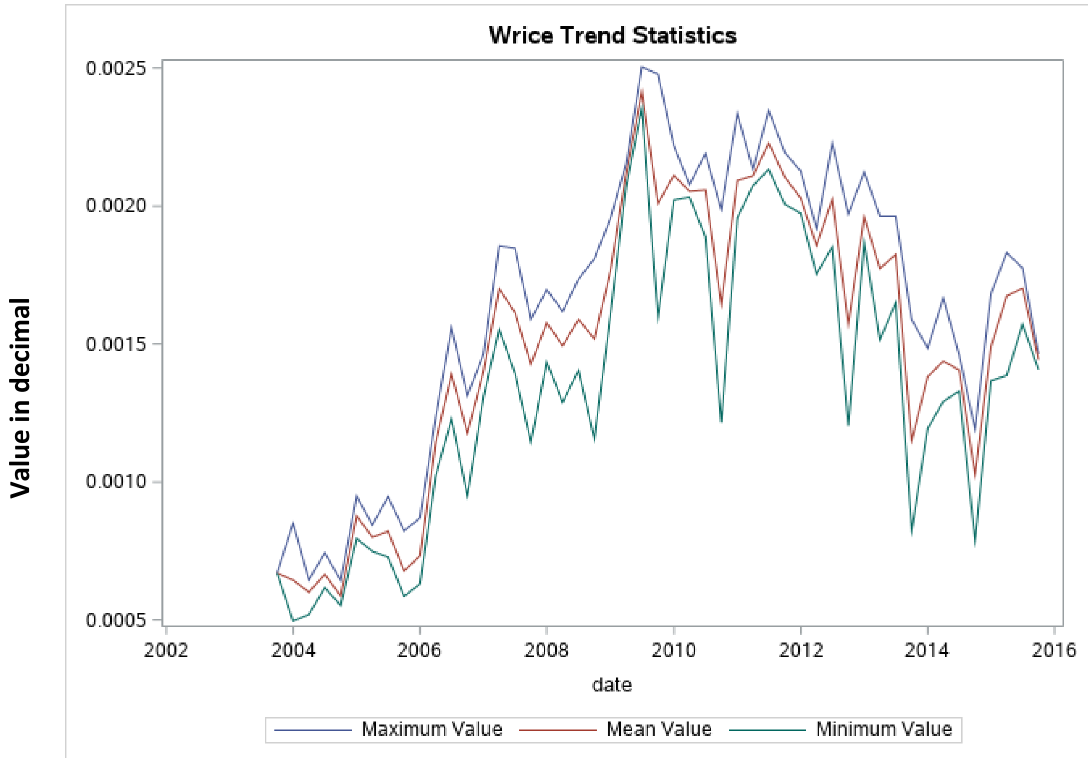


Figure 19 Trend Statistics in Budget Share: Rice Milk (Wrice)

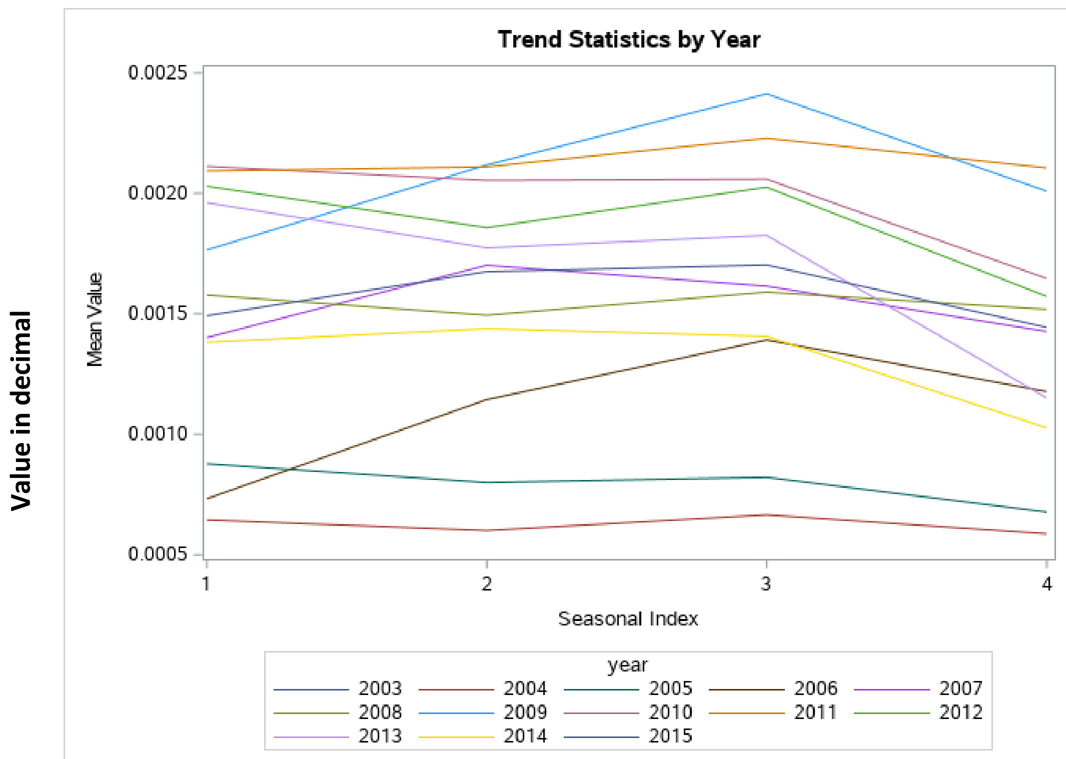


Figure 20 Trend Statistics by Year in Budget Share: Rice Milk (Wrice)

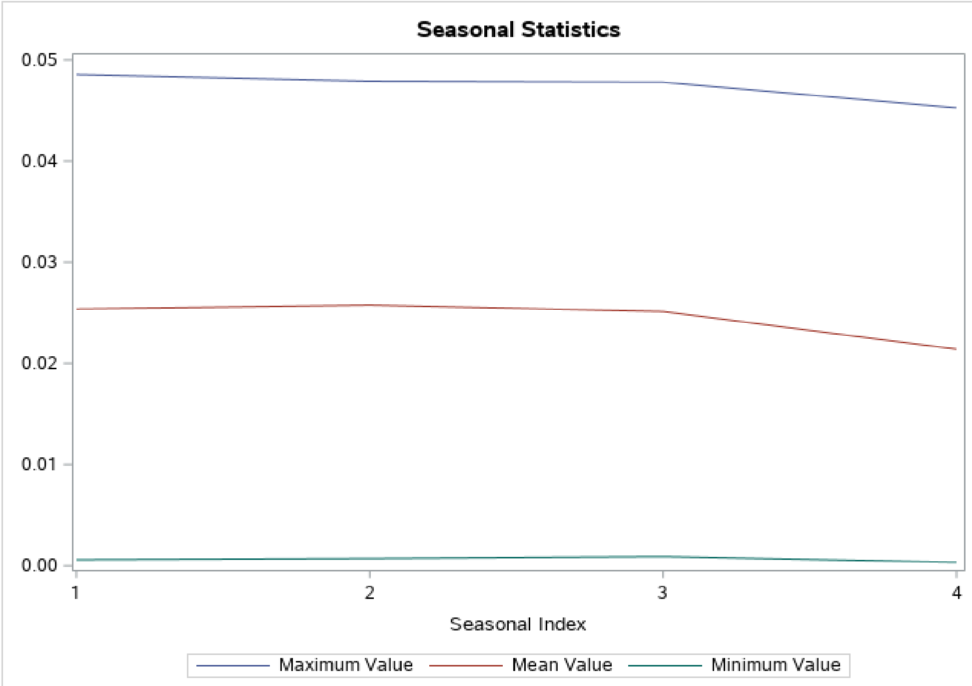


Figure 21 Seasonal Statistics in Budget Share: Rice Milk (Wrice)

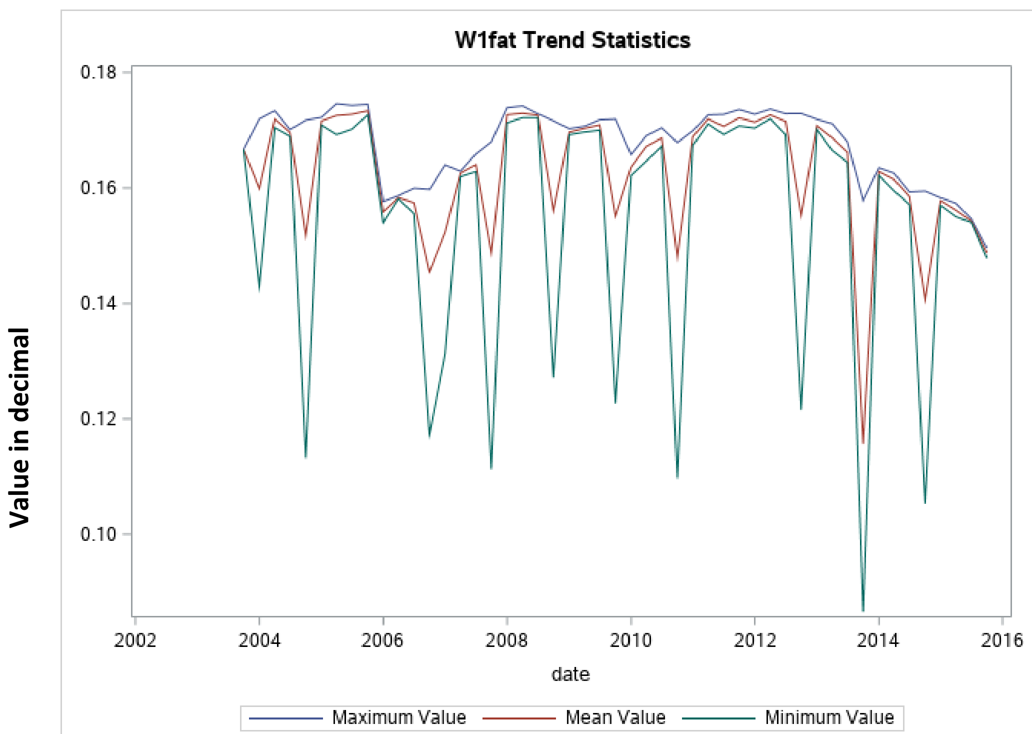


Figure 22 Trend Statistics in Budget Share: 1% Milk (W1fat)

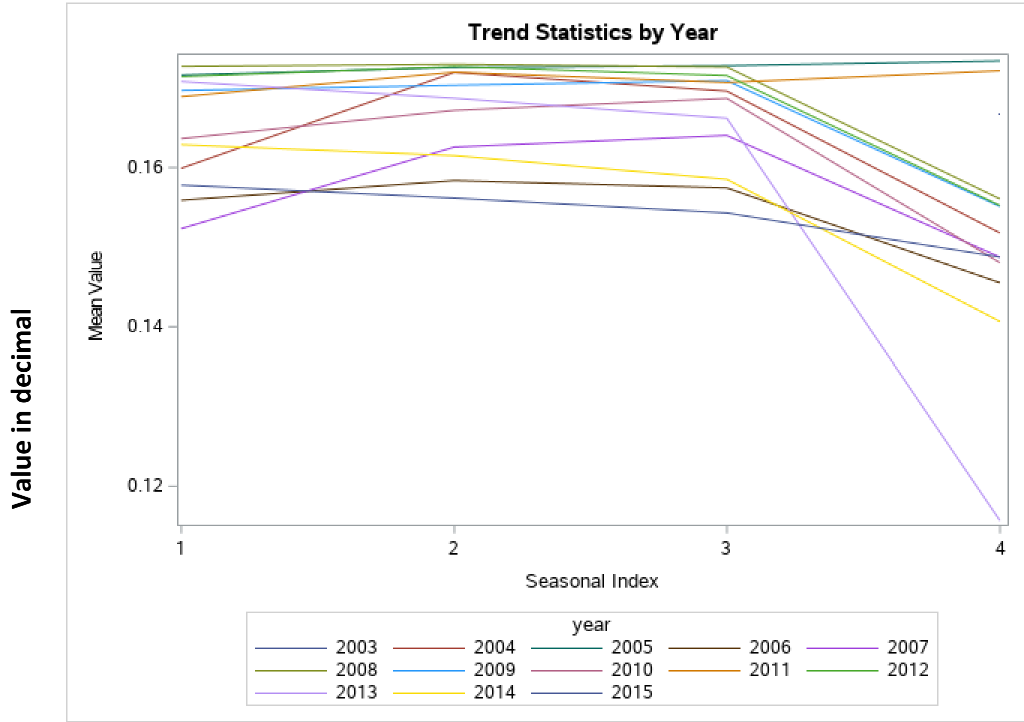


Figure 23 Trend Statistics by Year in Budget Share: 1% Milk (W1fat)

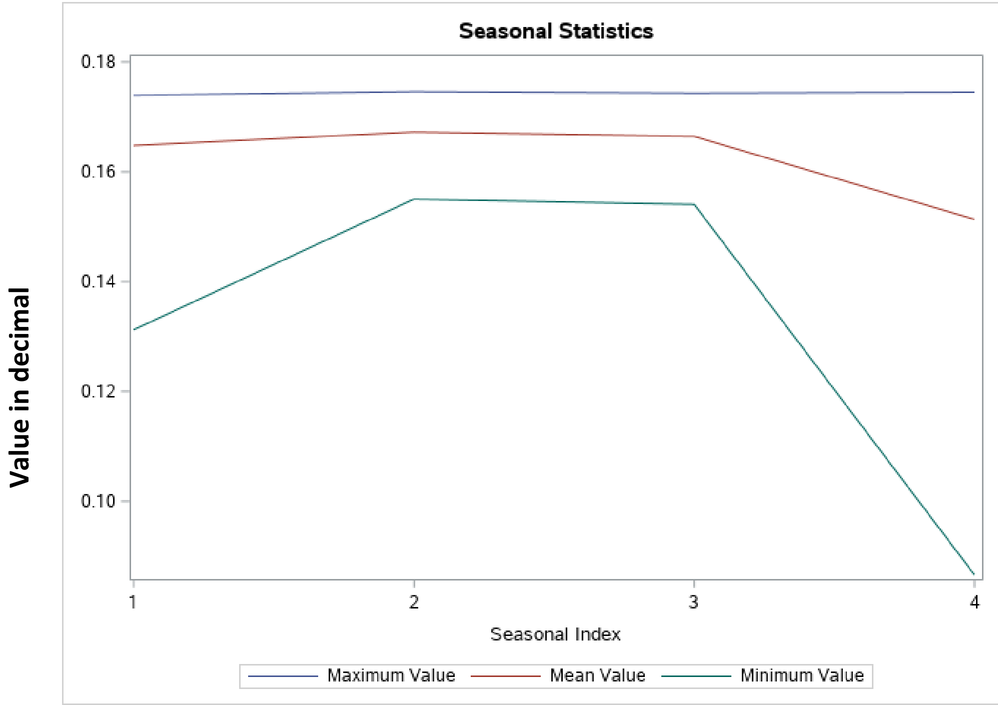


Figure 24 Seasonal Statistics in Budget Share: 1% Milk (W1fat)

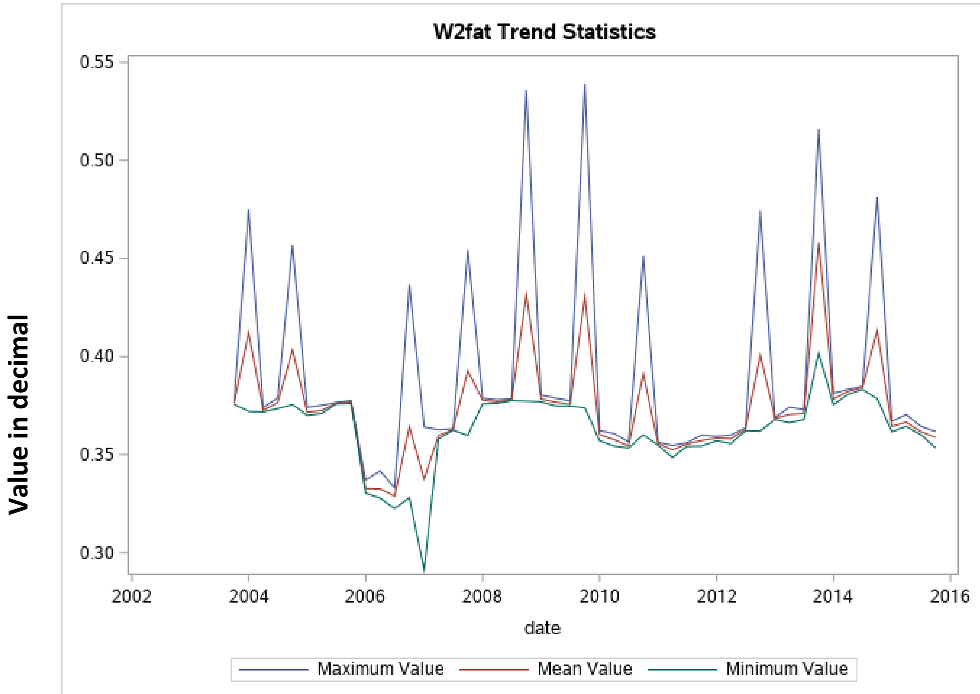


Figure 25 Trend Statistics in Budget Share: 2% Milk (W2fat)

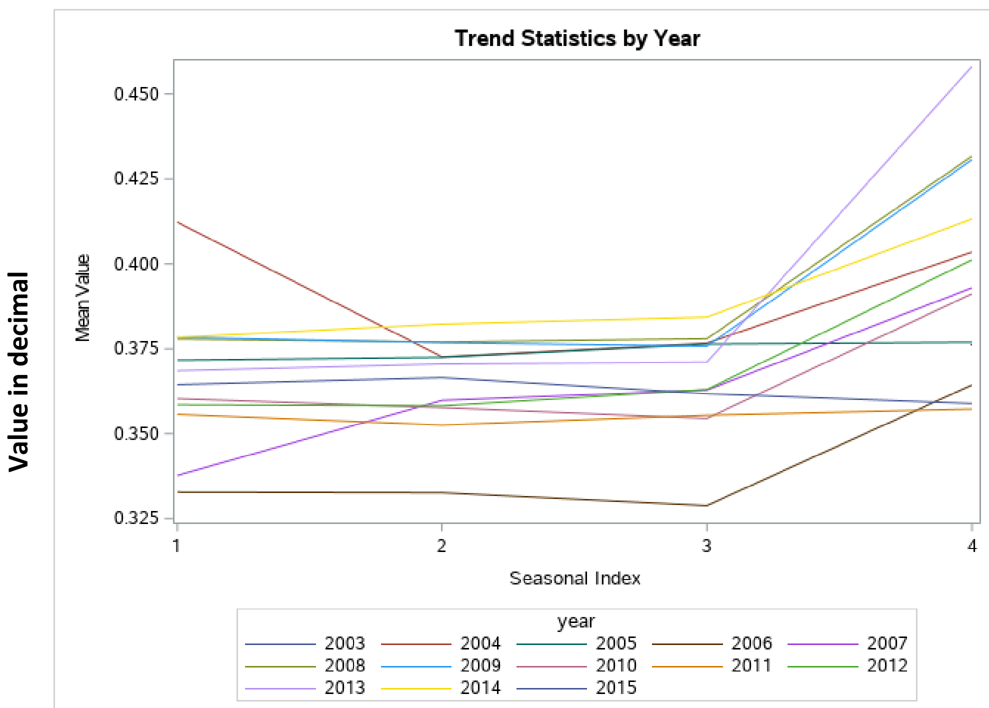


Figure 26 Trend Statistics by Year in Budget Share: 2% Milk (W2fat)

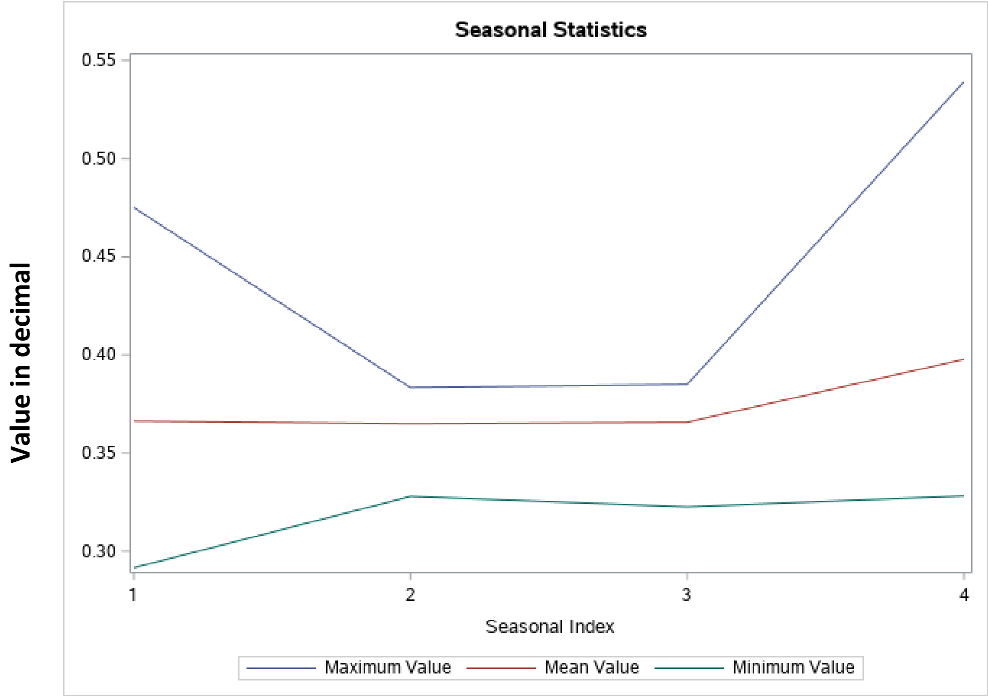


Figure 27 Seasonal Statistics in Budget Share: 2% Milk (W2fat)

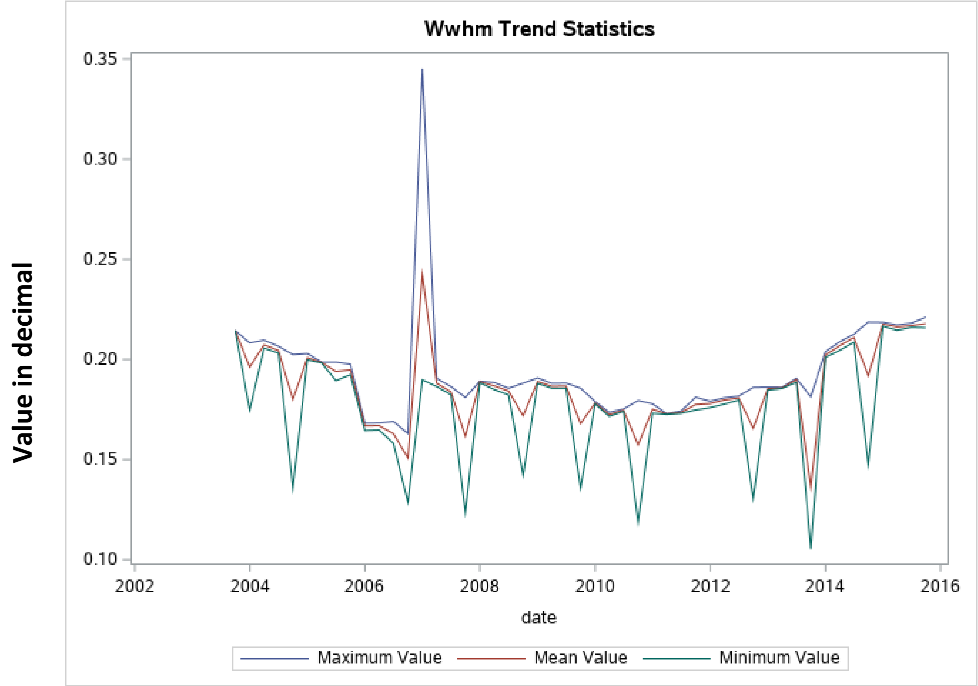


Figure 28 Trend Statistics in Budget Share: Whole Milk (Wwhm)

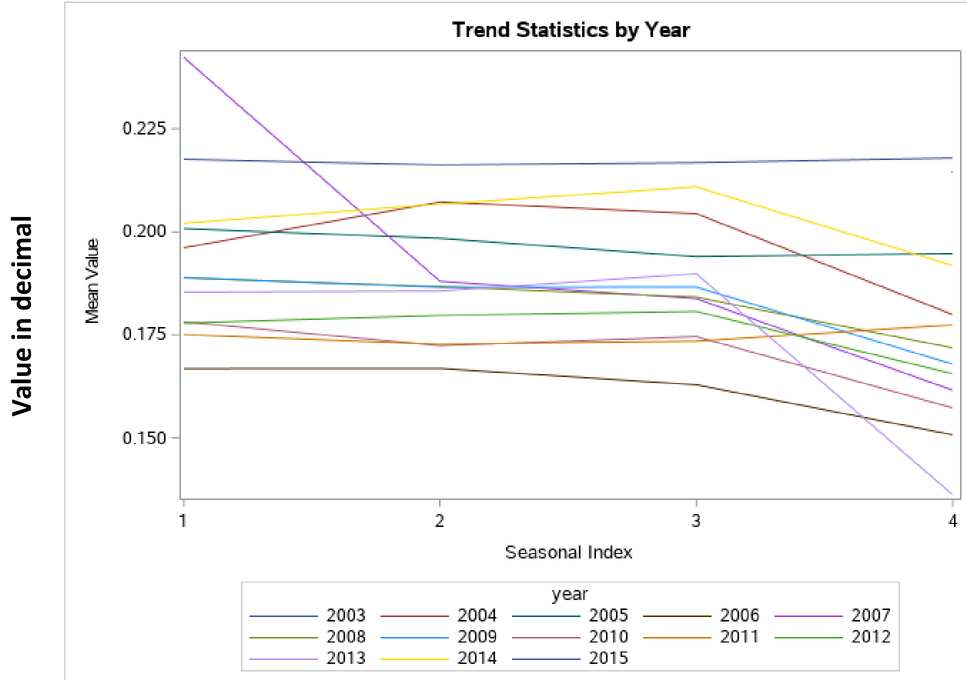


Figure 29 Trend Statistics by Year in Budget Share: Whole Milk (Wwhm)

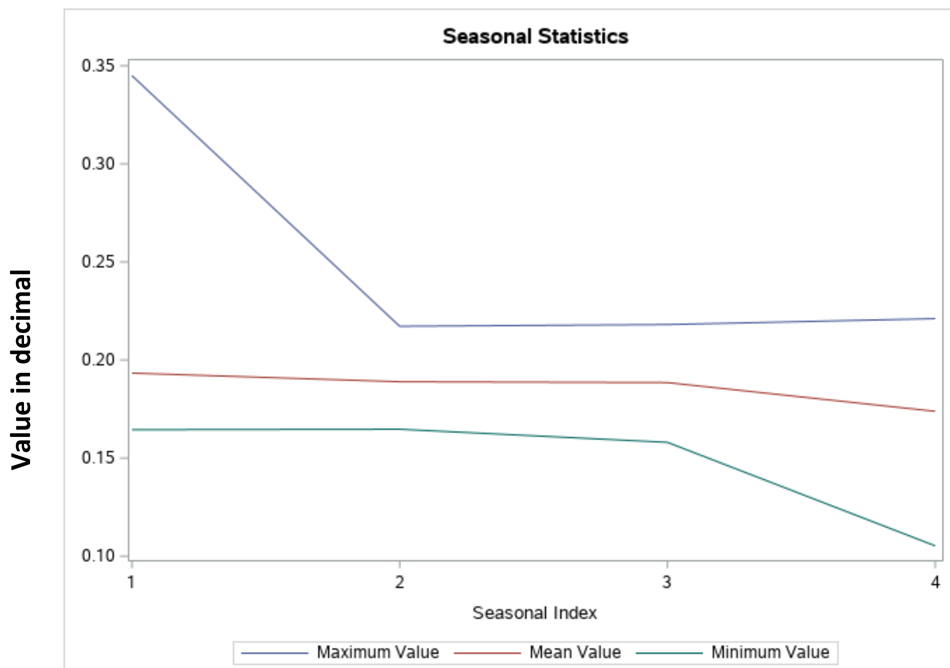


Figure 30 Seasonal Statistics in Budget Share: Whole Milk (Wwhm)

Figure 31 Trend Statistics in Budget Share: Fat-free Milk (Wffm)

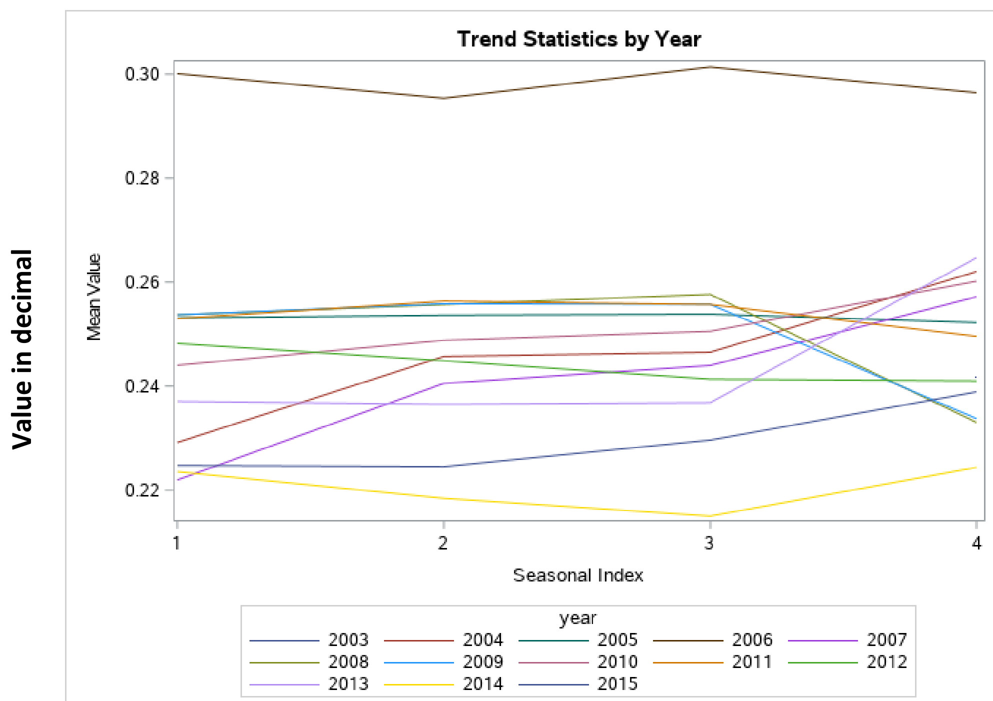


Figure 32 Trend Statistics by Year in Budget Share: Fat-free Milk (Wffm)

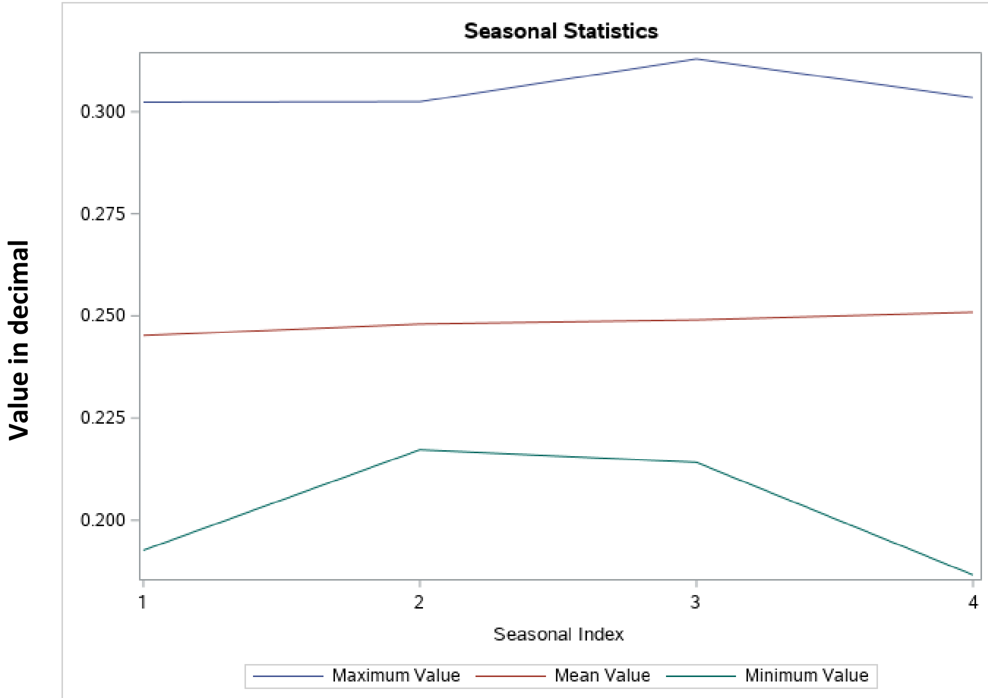


Figure 33 Seasonal Statistics in Budget Share: Fat-free Milk (Wffm)

The demand system is estimated using SAS 9.4 statistical software. We use the Proc Model procedure to estimate model parameters and subsequently to calculate expenditure, own-price and cross-price elasticities.

Model Estimates

We present the parameter estimates of the Barten synthetic model in table 28, table 32, table 36, and table 40. We separate the estimated results ino two major parts. The first part is estimation results of linear hedonic metric approached Barten Synthetic model and the second part is estimation results of log hedonic metric approached Barten Synthetic model. In each part, we tried to use both classic closeness index and exponential closeness index to calculate the own closeness index which is required to reparameterize the Barten Synthetic Model.

Linear Hedonic Matrix Augmented Barten Synthetic Model

In this section, we discuss the results of the estimation results of Barten Synthetic model based on linear hedonic matrix estimates. We estimated all models with no restrictions imposed but later we tested for symmetry and homogeneity. Also starting values are provided to aid convergence. Table 30 shows that there are 23 out of 37 parameters estimated were significant at p-value 0.05. The parameter estimates a_0, a_2 and c^h, c^{nn} are significant. Estimated Barten synthetic model was corrected for serial correlation using an *AR(5)* process. Table 30 also shows that calculated autocorrelation coefficients were statistically significant at 99% level indicating the presence of *AR(5)* disturbance terms. The joint hypotheses test for seasonal dummies, λ (lambda) and μ (mu) are shown in table 31. Significance (at 0.01 level) of seasonal (quarterly) dummy variables for almond milk, soy milk and fat-free milk confirms the presence of quarterly seasonality in the data set. The test of homogeneity in table 31 shows that it fails to reject six out of seven homogeneity restrictions. However, the symmetry test demonstrates mixed results. Moreover, the joint hypothesis for λ (lambda) and μ (mu) is rejected for possibility of data support for Rotterdam, AIDS, NBR and CBS versions of differential demand systems. This means that Barten Sythentic model itself is an appropriate demand model to model this data.

Based on the parameter estimates, we calculated uncompensated and compensated and expenditure elasticities for seven products considered in this study. Three out of seven budget share series were non-stationary; therefore, the sample mean over 145 observations are both candidates of local coordinates to evaluate elasticities. Therefore, we tried to obtain expenditure elasticities using both average of all the observations of each expenditure share and compare as well as average of the final twelve observations of each expenditure share to make a comparison.

The average of all observations are as follows: almond milk 0.003; soy milk 0.0244; rice milk 0.002; 2% milk 0.374; 1% milk 0.162; fat-free milk 0.248; whole milk 0.186; while the average of the last twelve observations are as follows, almond milk 0.004; soy milk 0.018; rice milk 0.001; 2% milk 0.453; 1% milk 0.126; fat-free milk 0.251; whole milk 0.147. We used elasticity formula showed in equations in chapter 2 to generate compensated and uncompensated own- and cross-price and expenditure elasticities respectively. Uncompensated own- and cross-price elasticities were generated using the derivative of Slutsky equation expressed in elasticity form. Table 32 shows the expenditure elasticities and uncompensated own-price and cross-price elasticities calculated using the final 12 observations of budget share respectively. It is shown that the calculated expenditure elasticity estimates except for rice milk are significant at or above the p-value 0.05 level. Soy milk is found to be the most expenditure elastic dairy milk alternative beverage. Because expenditure elasticity is a measure of the responsiveness of expenditure on, or consumption of, a good to a change in real income, this result also indicates that soy milk is the most responsive dairy milk alternative beverage for varying total expenditure values. Expenditure elasticities with respect to almond milk and soymilk are high due to the fact that the expenditure shares are low (expenditure shares are in the denominator of the expenditure elasticity calculation and that is why the expenditure elasticity numbers for almond milk and soymilk are somewhat high). Almond milk is also an expenditure elastic beverage (expenditure elasticity 3.5965). In terms of milk products, they are all expenditure inelastic (2% low-fat milk 0.8337; fat-free milk 0.5686; whole milk 0.5500) except for 1% low-fat milk (1.1392).

All uncompensated and compensated own-price elasticity estimates have negative sign. This result is indicative of theoretically consistent own-price elasticity estimates with demand theory, and they are statistically significant except for few exceptions. Compensated own-price elasticity of demand for soymilk is -0.2506, which is the highest. Small budget share and high prices associated with soymilk may have contributed to higher own-price elasticity of demand for soymilk. In another words, marginal consumers are more sensitive to a price change in soy milk compared to that of other dairy milk alternative beverages and milk products. All the milk alternative beverages under consideration showed inelastic own-price elasticity demands. Among the significant compensated own price elasticities, 2% low fat milk have the most inelastic elasticity of demand, which is -0.1.69. 17 out of 42 (40 percent) compensated cross-price elasticities have negative sign indicating net complements.

There is no significant substitute for almond milk. Soy milk's net substitutes are all four types of regular milk products. Soy milk is strong net complements for almond milk and rice milk. One possible explanation is that because the data we use is purchase data, it means that almond milk and soy milk are normally purchased together. All four types of regular milk products are strong substitutes between each other.

Table 34 illustrates the parameter estimates of Barten synthetic model using exponential closeness index. Similar with results of table 30, 22 out of 37 parameters are significant. Table 37 shows the compensated elasticities calculated using exponential own close index in the model. 16 out of 64 compensated cross-price elasticities have negative sign indicating net complements. 2% fat milk are found to be strong net substitute for almond milk and almond is net substitute of 2% low-fat milk, while soy milk, rice milk and 1% milk are net complements for almond milk. Similar with previous results, all four types of regular milk are net substitutes of soy milk. No strong net substitutes for rice milk. 2% milk can be substituted by soy milk and other three types of dairy milk products. Net substitutes for 1% milk are soy milk and other three types of dairy milk products. Soy milk is also strong substitute for fat-free milk and whole milk.

Table 30 Parameter Estimates of Linear Hedonic-Barten Synthetic Model for U.S. DABs consumed at Home: January 2004-December 2015

Parameter	Estimate	Std Err	P-value
a0	0.0089	0.0026	0.0007
a1	-0.0055	0.0032	0.0945
a2	-0.0052	0.0015	0.0008
ch	-0.0003	0.0001	<.0001
cnn	0.0003	0.0001	0.0034
b1	0.0081	0.0038	0.0341
lambda	1.2004	0.2694	<.0001
mu	0.1633	0.0199	<.0001
rho1	-0.7125	0.0360	<.0001
rho2	-0.5085	0.0408	<.0001
rho3	-0.3231	0.0425	<.0001
rho4	-0.1814	0.0417	<.0001
rho5	-0.1224	0.0340	0.0004
d11	0.0004	0.0001	0.0042
d12	0.0001	0.0001	0.7022
d13	0.0002	0.0001	0.0775
b2	0.2159	0.0309	<.0001
d21	0.0010	0.0013	0.4651
d22	0.0033	0.0014	0.0150
d23	0.0034	0.0014	0.0119
b3	0.0019	0.0038	0.6075
d31	-0.0002	0.0001	0.2247
d32	-0.0001	0.0001	0.3621
d33	-0.0003	0.0001	0.0307
b4	-0.1348	0.0988	0.1749
d41	-0.0008	0.0010	0.4394
d42	-0.0015	0.0010	0.1403
d43	-0.0016	0.0010	0.1164
b5	-0.0123	0.0451	0.7848
d51	-0.0013	0.0005	0.0175
d52	-0.0015	0.0005	0.0046
d53	-0.0015	0.0005	0.0042
b6	-0.1566	0.0721	0.0315
d61	0.0002	0.0007	0.7346
d62	-0.0005	0.0007	0.5151
d63	-0.0004	0.0007	0.5888
b7	-0.1226	0.0545	0.0261

Table 31 Joint Hypothesis Tests for Seasonal (Quarterly) Dummies and Lambda in Linear-Approached Barten Synthetic Model

Test Results				
Test0	Wald	10.56	0.0144	d11= d12= d13
Test1	Wald	9.19	0.0268	d21= d22= d23
Test2	Wald	5.87	0.1183	d31= d32= d33
Test3	Wald	3.91	0.2715	d41= d42= d43
Test4	Wald	16.88	0.0007	d51= d52= d53
Test5	Wald	0.87	0.8331	d61= d62= d63
Test6	Wald	81.86	<.0001	lambda=0,mu=0
Test7	Wald	1787.5	<.0001	lambda=1,mu=1
Test8	Wald	67.41	<.0001	lambda=1,mu=0
Test9	Wald	1832.4	<.0001	lambda=0,mu=1
Test10	Wald	5.52	0.0188	g11 + g12 + g13 + g14 + g15 + g16 + g17 = 0
Test11	Wald	13.25	0.0003	g21 + g22 + g23 + g24 + g25 + g26 + g27 = 0
Test12	Wald	2.41	0.1205	g31 + g32 + g33 + g34 + g35 + g36 + g37 = 0
Test13	Wald	0.38	0.5375	g41 + g42 + g43 + g44 + g45 + g46 + g47 = 0
Test14	Wald	0.11	0.7384	g51 + g52 + g53 + g54 + g55 + g56 + g57 = 0
Test15	Wald	0.79	0.3741	g61 + g62 + g63 + g64 + g65 + g66 + g67 = 0
Test16	Wald	1.30	0.2534	g12=g21
Test17	Wald	16.61	<.0001	g13=g31
Test18	Wald	16.61	<.0001	g14=g41
Test19	Wald	16.61	<.0001	g15=g51
Test20	Wald	16.61	<.0001	g16=g61
Test21	Wald	16.61	<.0001	g23=g32
Test22	Wald	16.61	<.0001	g24=g42
Test23	Wald	17.44	<.0001	g25=g52
Test24	Wald	16.61	<.0001	g26=g62
Test25	Wald	16.61	<.0001	g34=g43
Test26	Wald	16.61	<.0001	g35=g53
Test27	Wald	0.81	0.3686	g36=g63
Test28	Wald	0.16	0.6936	g45=g54
Test29	Wald	16.61	<.0001	g46=g64
Test30	Wald	16.61	<.0001	g56=g65

Table 32 Expenditure Elasticities and Uncompensated Own-Price and Cross-Price Demand Elasticities Estimated through Linear Hedonic-Barten Synthetic Model

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk	expenditure
almond milk	-0.1315*	-0.0557	-0.0618***	-1.334***	-0.6149***	-0.9153***	-0.6800***	3.5965***
	0.0564	0.0446	0.0137	0.4062	0.1764	0.2705	0.2029	0.0014
soy milk	-0.0321***	-0.4966***	-0.0234***	-3.7103***	-1.6169***	-2.4583***	-1.8488***	10.0746***
	0.0056	0.0428	0.0027	0.4510	0.1958	0.2994	0.2245	1.2054
rice milk	-0.1281***	-0.1872***	-0.1019	-0.9056	-0.4662	-0.4956	-0.4920	2.3121
	0.0284	0.0578	0.1263	0.9263	0.3963	0.6155	0.4576	2.5043
2% milk	-0.0027***	-0.0167***	-0.0014***	-0.4186***	-0.1076***	-0.1653***	-0.1238***	0.8337***
	0.0003	0.0018	0.0002	0.0275	0.0115	-0.1653	0.0132	0.0680
1% milk	-0.0044***	-0.0250***	-0.0026***	-0.3619***	-0.3325***	-0.2419***	-0.1812***	1.1392***
	0.0004	0.0020	0.0003	0.0306	0.0222	0.0203	0.0152	0.0826
fat-free milk	-0.0022***	-0.0096***	-0.0004	-0.1498***	-0.0656***	-0.2777***	-0.0747***	0.5686***
	0.0003	0.0019	0.0004	0.0278	0.0121	0.0217	0.0138	0.0708
whole milk	-0.0020***	-0.0100***	-0.0013***	-0.1427***	-0.0625***	-0.0951***	-0.2246***	0.5500***
	0.0004	0.0024	0.0003	0.0365	0.0159	0.0243	0.0250	0.0958

Table 33 Compensated Own-Price and Cross-Price Demand Elasticities Estimated through Linear Hedonic-Barten Synthetic Model

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	-0.1193*	-0.1311***	-0.0563***	0.0108	-0.0308*	-0.0223	-0.0106
	0.0558	0.0349	0.0139	0.0144	0.0144	-0.0162	0.0107
soy milk	0.0022	-0.2506***	-0.0081***	0.0563***	0.0191***	0.0433***	0.0261***
	0.0041	0.0268	0.0021	0.0074	0.0037	0.0066	0.0038
rice milk	-0.1202***	-0.1307***	-0.0984	-0.0411	-0.0907***	0.0786	-0.0617***
	0.0304	0.0340	0.1264	0.0270	0.0298	0.0658	0.0236
2% milk	0.0304	0.0037***	-0.0002	-0.1069***	0.0278***	0.0417***	0.0313***
	0.0001	0.0005	0.0001	0.0122	0.0032	0.0048	0.0036
1% milk	-0.0006**	0.0029***	-0.0009***	0.0641***	-0.1475***	0.0410***	0.0308***
	0.0003	0.0006	0.0003	0.0073	0.0159	0.0048	0.0036
fat-free milk	-0.0002	0.0043***	0.0005	0.0629***	0.0268***	-0.1365***	0.0311***
	0.0002	0.0006	0.0004	0.0072	0.0032	0.0144	0.0036
whole milk	-0.0001	0.0034***	-0.0005***	0.0629***	0.0268***	0.0415***	-0.1222***
	0.0002	0.0005	0.0002	0.0072	0.0032	0.0048	0.0165

Table 34 Parameter Estimates of Linear Hedonic-Barten Synthetic model (with Exponential Closeness Index) for U.S. DABs Consumed at Home: January 2004-December 2015

Parameter	Estimate	Std Err	P-value
a0	0.0199	0.0065	0.0027
a1	-0.0052	0.0033	0.1187
a2	-0.0049	0.0016	0.0028
ch	-0.0003	0.0001	0.0001
cnn	0.0003	0.0001	0.0045
b1	0.0088	0.0038	0.0225
lambda	1.1922	0.2707	<.0001
mu	0.1649	0.0200	<.0001
rho1	-0.7110	0.0360	<.0001
rho2	-0.5075	0.0408	<.0001
rho3	-0.3226	0.0425	<.0001
rho4	-0.1776	0.0416	<.0001
rho5	-0.1195	0.0339	0.0006
d11	0.0004	0.0001	0.0057
d12	0.0000	0.0001	0.7587
d13	0.0002	0.0001	0.0898
b2	0.2125	0.0309	<.0001
d21	0.0010	0.0013	0.4727
d22	0.0033	0.0014	0.0162
d23	0.0034	0.0014	0.0131
b3	0.0021	0.0038	0.5771
d31	-0.0002	0.0001	0.2169
d32	-0.0001	0.0001	0.3537
d33	-0.0003	0.0001	0.0301
b4	-0.1302	0.0994	0.1924
d41	-0.0008	0.0010	0.4425
d42	-0.0015	0.0010	0.1421
d43	-0.0016	0.0010	0.1180
b5	-0.0105	0.0453	0.8179
d51	-0.0013	0.0005	0.0185
d52	-0.0015	0.0005	0.0050
d53	-0.0015	0.0005	0.0046
b6	-0.1543	0.0724	0.0348
d61	0.0003	0.0007	0.7106
d62	-0.0004	0.0007	0.5378
d63	-0.0004	0.0007	0.5378
b7	-0.1206	0.0547	0.0290

Table 35 Joint Hypothesis Tests for Seasonal (Quarterly) Dummies and Lambda in Linear-Barten Synthetic Model

Test Results				
Test0	Wald	9.86	0.0198	d11= d12= d13
Test1	Wald	8.97	0.0297	d21= d22= d23
Test2	Wald	5.96	0.1136	d31= d32= d33
Test3	Wald	3.87	0.2763	d41= d42= d43
Test4	Wald	16.61	0.0009	d51= d52= d53
Test5	Wald	0.83	0.8434	d61= d62= d63
Test6	Wald	82.00	<.0001	lambda=0,mu=0
Test7	Wald	1767.2	<.0001	lambda=1,mu=1
Test8	Wald	68.19	<.0001	lambda=1,mu=0
Test9	Wald	1812.8	<.0001	lambda=0,mu=1
Test10	Wald	6.70	0.0097	g11 + g12 + g13 + g14 + g15 + g16 + g17 = 0
Test11	Wald	10.80	0.0010	g21 + g22 + g23 + g24 + g25 + g26 + g27 = 0
Test12	Wald	2.42	0.1195	g31 + g32 + g33 + g34 + g35 + g36 + g37 = 0
Test13	Wald	0.89	0.3464	g41 + g42 + g43 + g44 + g45 + g46 + g47 = 0
Test14	Wald	0.04	0.8356	g51 + g52 + g53 + g54 + g55 + g56 + g57 = 0
Test15	Wald	0.76	0.3847	g61 + g62 + g63 + g64 + g65 + g66 + g67 = 0
Test16	Wald	1.16	0.2810	g12=g21
Test17	Wald	15.65	<.0001	g13=g31
Test18	Wald	15.65	<.0001	g14=g41
Test19	Wald	15.65	<.0001	g15=g51
Test20	Wald	15.65	<.0001	g16=g61
Test21	Wald	15.65	<.0001	g23=g32
Test22	Wald	15.65	<.0001	g24=g42
Test23	Wald	16.35	<.0001	g25=g52
Test24	Wald	15.65	<.0001	g26=g62
Test25	Wald	15.65	<.0001	g34=g43
Test26	Wald	15.65	<.0001	g35=g53
Test27	Wald	0.71	0.4006	g36=g63
Test28	Wald	0.12	0.7273	g45=g54
Test29	Wald	15.65	<.0001	g46=g64
Test30	Wald	15.65	<.0001	g56=g65

Table 36 Expenditure Elasticities and Uncompensated Own-Price and Cross-Price Demand Elasticities Estimated through Linear Hedonic-Barten Synthetic Model (with Exponential Closeness Index)

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk	expenditure
almond milk	-0.1667***	-0.0358	-0.0546***	-1.5942***	-0.4812***	-0.9159***	-0.5393***	3.5843***
	0.0517	0.0368	0.0127	0.4595	0.1263	0.2538	0.1482	1.0204
soy milk	-0.0416***	-0.4933***	-0.0258***	-5.7855***	-1.6139***	-3.1999***	-1.8808***	12.9127***
	0.0078	0.0501	0.0034	0.7492	0.2078	0.4152	0.2430	1.6521
rice milk	-0.1769***	-0.2383***	0.088	-1.4593	-0.5276	-0.6294	-0.5551	3.0680
	0.0392	0.0615	0.1816	1.5006	0.4066	0.8334	0.4800	3.3418
2% milk	-0.0031***	-0.0138***	-0.0012***	-0.4986***	-0.0927***	-0.1864***	-0.1090***	0.9049***
	0.0003	0.0015	0.0001	0.0380	0.0103	0.0207	0.0121	0.0804
1% milk	-0.0050***	-0.0188***	-0.0025***	-0.4266***	-0.2905***	-0.2390***	-0.1399***	1.1091***
	0.0006	0.0024	0.0004	0.0585	0.0245	0.0324	0.0189	0.1296
fat-free milk	-0.0023***	-0.0072***	0.0000***	-0.0942***	-0.0528***	-0.2775***	-0.0613***	0.5784***
	0.0003	0.0014	0.0004	0.0668	0.0092	0.0218	0.0107	0.0697
whole milk	-0.0017***	-0.0047*	-0.0012***	-0.1880	-0.0269	-0.0527	-0.1580***	0.3710**
	0.0006	0.0027	0.0003	0.0330	0.0185	0.0371	0.0316	0.1455

Table 37 Compensated Own-Price and Cross-Price Demand Elasticities Estimated through Linear Hedonic-Barten Synthetic Model (with Exponential Closeness Index)

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	-0.1535***	-0.1357***	-0.0505***	0.0297**	-0.0307**	-0.0149	-0.0125
	0.0512	0.0335	0.0128	0.0145	0.0132	0.0151	0.0097
soy milk	0.0059	-0.2592***	-0.0113***	0.0648***	0.0090**	0.0461***	0.0167***
	0.0056	0.0344	0.0029	0.0094	0.0039	0.0077	0.0035
rice milk	-0.1656***	-0.1826***	0.0921	-0.0693*	-0.1420***	0.1419	-0.1042***
	0.0420	0.0469	0.1816	0.0375	0.0412	0.0908	0.0326
2% milk	0.0002**	0.0026***	-0.0002*	-0.0887***	0.0210***	0.0411***	0.0240***
	0.0001	0.0004	0.0001	0.0111	0.0025	0.0050	0.0029
1% milk	-0.0009**	0.0013**	-0.0013***	0.0758***	-0.1511***	0.0398***	0.0231***
	0.0004	0.0006	0.0004	0.0091	0.0168	0.0050	0.0030
fat-free milk	-0.0002	0.0033***	0.0006	0.0740***	0.0199***	-0.1321***	0.0237***
	0.0002	0.0006	0.0004	0.0091	0.0025	0.0144	0.0029
whole milk	-0.0003	0.0021***	-0.0008***	0.0739***	0.0198***	0.0406***	-0.1034***
	0.0002	0.0004	0.0002	0.0091	0.0025	0.0050	0.0230

Log Hedonic Matrix Augmented Barten Synthetic Model

Table 38 shows the results of Baten synthetic model results using log hedonic matrix estimates. 20 out of 36 estimated parameters were significant at the p-value 0.05. According to table 38, the parameter estimates a_0, a_1, a_2 are significant but c^h, c^{nn} are not significant. Estimated Barten synthetic model was corrected for serial correlation using an *AR(5)* process of disturbance terms. Table 38 shows that calculated autocorrelation coefficients were statistically significant at 99% level indicating the presence of *AR(5)* disturbance terms. The joint hypotheses test for seasonal dummies, λ (lambda) and μ (mu) are shown in table 39. Significance (at 0.10 level) of seasonal (quarterly) dummy Significance (at 0.01 level) of seasonal (quarterly) dummy variables for almond milk, soy milk and 1% fat milk confirms the presence of quarterly seasonality in the data set. Moreover, the joint hypothesis for λ (lambda) and μ (mu) is rejected for possibility of data support for Rotterdam, AIDS, NBR and CBS versions of differential demand systems. This means that Barten Sythentic model itself is a valid demand system to model this data.

Table 40 and 44 present the expenditure elasticities and uncompensated own-price and cross-price elasticities calculated using classic own closeness index and exponential own closeness index respectively. Similarly, we see that the calculated expenditure elasticity estimates except for rice milk are significant at p-value 0.05 level. Soy milk is also found to be the most expenditure elastic dairy alternative beverage.

19 out of 42 (45%) compensated cross-price elasticities have negative sign and 55% of compensated cross-price elasticities have positive sign. There is no strong substitute for almond milk. This argument applies to rice milk as well. Soy milk and other types of dairy milk products

are found to be substitutes for 2% low-fat milk which is the same as the results we summarized above. Also soy milk is a substitute for fat-free milk and whole milk. As shown in table 45, 21 out of 42 (50%) compensated cross-price elasticities show complementary behavior while another half of cross-price elasticities exhibit substitution effects. Soy milk and rice milk act as net complements to almond milk. Soy milk is continuously to be a substitute for all conventional milk types. Almond and soy milk and four dairy milk products are all complements of rice milk. The same strong substitution effects can be found among all four dairy milk products.

To conclude, from both linear hedonic augmented and log hedonic augmented Barten synthetic model, we observe strong substitution effects between soy milk and four types of dairy milk products. Moreover, we find almond milk and 2% low-fat milk are substitutes. Even though the results do not indicate that rice milk can be a substitute for dairy milk, it serves as complements for soy milk and almond milk.

Table 38 Parameter Estimates of Log Hedonic-Barten Synthetic Model for U.S. DABs Consumed at Home: January 2004-December 2015

Parameter	Estimate	Std Err	P-value
a0	0.0088	0.0038	0.0224
a1	-0.0142	0.0063	0.0261
a2	-0.0061	0.0027	0.0262
ch	-0.0004	0.0001	0.0093
cnn	-0.0004	0.0002	0.1061
b1	0.0119	0.0037	0.0017
lambda	1.3137	0.2672	<.0001
mu	0.1715	0.0195	<.0001
rho1	-0.6874	0.0356	<.0001
rho2	-0.5013	0.0409	<.0001
rho3	-0.3393	0.0429	<.0001
rho4	-0.1904	0.0427	<.0001
rho5	-0.1128	0.0344	0.0013
d11	0.0003	0.0001	0.0201
d12	0.0000	0.0001	0.9451
d13	0.0002	0.0001	0.1974
b2	0.1971	0.0303	<.0001
d21	0.0009	0.0013	0.4833
d22	0.0032	0.0013	0.0174
d23	0.0033	0.0013	0.0142
b3	0.0025	0.0038	0.5099
d31	-0.0002	0.0001	0.1759
d32	-0.0001	0.0001	0.3136
d33	-0.0003	0.0001	0.0263
b4	-0.1946	0.0980	0.0492
d41	-0.0004	0.0010	0.7174
d42	-0.0011	0.0010	0.2932
d43	-0.0012	0.0010	0.2544
b5	-0.0216	0.0454	0.6353
d51	-0.0014	0.0006	0.0155
d52	-0.0016	0.0006	0.0050
d53	-0.0016	0.0006	0.0048
b6	-0.1673	0.0718	0.0213
d61	0.0000	0.0007	0.9882
d62	-0.0007	0.0007	0.3202
d63	-0.0006	0.0007	0.3773
b7	-0.1418	0.0542	0.0099

Table 39 Joint Hypothesis Tests for Seasonal (Quarterly) Dummies and Lambda in Log-Approached Barten Synthetic Model

Test Results				
Test0	Wald	6.89	0.0754	d11= d12= d13
Test1	Wald	8.78	0.0323	d21= d22=d23
Test2	Wald	6.50	0.0897	d31= d32= d33
Test3	Wald	1.98	0.5760	d41= d42=d43
Test4	Wald	16.59	0.0009	d51=d52=d53
Test5	Wald	1.51	0.6793	d61=d62=d63
Test6	Wald	99.40	<.0001	lambda=0,mu=0
Test7	Wald	1816.4	<.0001	lambda=1,mu=1
Test8	Wald	78.43	<.0001	lambda=1,mu=0
Test9	Wald	1848.1	<.0001	lambda=0,mu=1
Test10	Wald	3.50	0.0612	g11 + g12 + g13 + g14 + g15 + g16 + g17 = 0
Test11	Wald	7.48	0.0062	g21 + g22 + g23 + g24 + g25 + g26 + g27 = 0
Test12	Wald	0.87	0.3519	g31 + g32 + g33 + g34 + g35 + g36 + g37 = 0
Test13	Wald	4.58	0.0323	g41 + g42 + g43 + g44 + g45 + g46 + g47 = 0
Test14	Wald	2.34	0.1264	g51 + g52 + g53 + g54 + g55 + g56 + g57 = 0
Test15	Wald	1.12	0.2908	g61 + g62 + g63 + g64 + g65 + g66 + g67 = 0
Test16	Wald	6.89	0.0087	g12=g21
Test17	Wald	6.96	0.0083	g13=g31
Test18	Wald	6.96	0.0083	g14=g41
Test19	Wald	6.96	0.0083	g15=g51
Test20	Wald	6.96	0.0083	g16=g61
Test21	Wald	5.62	0.0178	g23=g32
Test22	Wald	3.60	0.0580	g24=g42
Test23	Wald	6.96	0.0083	g25=g52
Test24	Wald	6.96	0.0083	g26=g62
Test25	Wald	6.96	0.0083	g34=g43
Test26	Wald	6.96	0.0083	g35=g53
Test27	Wald	0.13	0.7203	g36=g63
Test28	Wald	6.96	0.0083	g45=g54
Test29	Wald	6.96	0.0083	g46=g64
Test30	Wald	6.96	0.0083	g56=g65

Table 40 Expenditure Elasticities and Uncompensated Own-Price and Cross-Price Demand Elasticities Estimated through Log Hedonic-Barten Synthetic Model

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk	expenditure
almond milk	-0.0825	-0.1646***	-0.0756***	-2.0514***	-0.6064***	-1.1480***	-0.7036***	4.5544***
	0.0518	0.0367	0.0268	0.4497	0.1261	0.2497	0.1471	0.9940
soy milk	-0.0615***	-0.4470***	-0.0296***	-5.4585***	-1.5212***	-3.0309***	-1.7784***	12.1875***
	0.0086	0.0487	0.0063	0.7371	0.2044	0.4090	0.2389	1.6265
rice milk	-0.2438***	-0.3216***	0.0914	-1.7297	-0.5878	-1.0132	-0.6941	3.5189
	0.0859	0.1037	0.1621	1.4885	0.4079	0.8227	0.4768	3.3215
2% milk	-0.0032***	-0.0135***	-0.0013***	-0.5032***	-0.0900***	-0.1796***	-0.1052***	0.8843***
	0.0004	-0.0135	0.0002	0.0377	0.0106	0.0211	0.0123	0.0818
1% milk	-0.0052***	-0.0192***	-0.0026***	-0.4412***	-0.2865***	-0.2452***	-0.1471***	1.1420***
	0.0007	0.0024	0.0006	0.0610	0.0255	0.0338	0.0196	0.1354
fat-free milk	-0.0024***	-0.0094***	-0.0013***	-0.2167***	-0.0605***	-0.2894***	-0.0707***	0.6481***
	0.0004	0.0013	0.0003	0.0322	0.0089	0.0212	0.0105	0.0669
whole milk	-0.0021***	-0.0048*	-0.0017***	-0.0817	-0.0261	-0.0457	-0.2022***	0.3487**
	0.0008	0.0027	0.0006	0.0664	0.0185	0.0368	0.0275	0.1451

Table 41 Compensated Own-Price and Cross-Price Demand Elasticities Estimated through Log Hedonic-Barten Synthetic Model

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	-0.0658	-0.2536***	-0.0704***	0.0120	-0.0340	-0.0031	-0.0343
	0.0508	0.0374	0.0268	0.0265	0.0212	0.0182	0.0228
soy milk	-0.0167**	-0.2261***	-0.0160***	0.0631***	0.0106**	0.0329***	0.0126**
	-0.0167	0.0338	0.0061	0.0104	0.0048	0.0063	0.0056
rice milk	-0.2308***	-0.2578***	0.0954	-0.1355*	-0.1455**	-0.1286**	-0.1770**
	0.0878	0.0989	0.1620	0.0813	0.0634	0.0653	0.0767
2% milk	0.0001	0.0025**	-0.0003*	-0.1026***	0.0211***	0.0427***	0.0248***
	0.0002	0.0004	0.0002	0.0110	0.0025	0.0049	0.0029
1% milk	-0.0010	0.0015	-0.0013**	0.0762***	-0.1430***	0.0419***	0.0207***
	0.0006	0.0007	0.0006	0.0088	0.0176	0.0049	0.0037
fat-free milk	-0.0001	0.0024***	-0.0006**	0.0770***	0.0210***	-0.1265***	0.0246***
	0.0003	0.0005	0.0003	0.0088	0.0025	0.0148	0.0029
Whole milk	-0.0009	0.0016**	-0.0014**	0.0763***	0.0177***	0.0420***	-0.1510***
	0.000419	0.000622	0.000431	0.00701	0.00331	0.00466	0.0152

Table 42 Parameter Estimates of Log Hedonic-Barten Synthetic model (with Exponential Closeness Index)

Parameter	Estimate	Std Err	P-value
b1	-0.0281	0.0181	0.1226
lambda	-0.0113	0.0069	0.1032
a0	0.0071	0.0046	0.1195
a1	-0.0004	0.0001	0.0126
a2	-0.0003	0.0002	0.1251
mu	0.0127	0.0037	0.0009
ch	1.2912	0.2677	<.0001
cnn	0.1729	0.0195	<.0001
rho1	-0.6870	0.0356	<.0001
rho2	-0.5007	0.0409	<.0001
rho3	-0.3384	0.0429	<.0001
rho4	-0.1851	0.0425	<.0001
d11	-0.1085	0.0344	0.002
d12	0.0003	0.0001	0.0274
d13	0.0000	0.0001	0.8824
b2	0.0002	0.0001	0.2245
d21	0.1926	0.0303	<.0001
d22	0.0009	0.0013	0.4779
d23	0.0032	0.0013	0.0182
b3	0.0033	0.0013	0.0152
d31	0.0026	0.0037	0.485
d32	-0.0002	0.0001	0.1714
d33	-0.0001	0.0001	0.3086
b4	-0.0003	0.0001	0.0263
d41	-0.1831	0.0984	0.0649
d42	-0.0004	0.0010	0.6996
d43	-0.0011	0.0010	0.2812
b5	-0.0012	0.0010	0.2434
d51	-0.0185	0.0454	0.684
d52	-0.0014	0.0006	0.0175
d53	-0.0016	0.0006	0.0058
b6	-0.0016	0.0006	0.0055
d61	-0.1608	0.0719	0.027
d62	0.0000	0.0007	0.9891
d63	-0.0007	0.0007	0.3381
b7	-0.13673	0.0543	0.0130

Table 43 Joint Hypothesis Tests for Seasonal (Quarterly) Dummies and Lambda in Log-Barten Synthetic Model (with Exponential Closeness Index)

Test Results				
Test0	Wald	6.26	0.0996	d11= d12= d13
Test1	Wald	8.62	0.0348	d21= d22= d23
Test2	Wald	6.55	0.0878	d31= d32= d33
Test3	Wald	2.08	0.5559	d41= d42= d43
Test4	Wald	16.05	0.0011	d51= d52= d53
Test5	Wald	1.41	0.7020	d61= d62= d63
Test6	Wald	99.93	<.0001	lambda=0,mu=0
Test7	Wald	1810.1	<.0001	lambda=1,mu=1
Test8	Wald	79.59	<.0001	lambda=1,mu=0
Test9	Wald	1840.7	<.0001	lambda=0,mu=1
Test10	Wald	4.10	0.0429	g11 + g12 + g13 + g14 + g15 + g16 + g17 = 0
Test11	Wald	5.21	0.0224	g21 + g22 + g23 + g24 + g25 + g26 + g27 = 0
Test12	Wald	0.71	0.3994	g31 + g32 + g33 + g34 + g35 + g36 + g37 = 0
Test13	Wald	2.52	0.1124	g41 + g42 + g43 + g44 + g45 + g46 + g47 = 0
Test14	Wald	1.31	0.2518	g51 + g52 + g53 + g54 + g55 + g56 + g57 = 0
Test15	Wald	1.20	0.2729	g61 + g62 + g63 + g64 + g65 + g66 + g67 = 0
Test16	Wald	6.45	0.0111	g12=g21
Test17	Wald	6.39	0.0115	g13=g31
Test18	Wald	6.39	0.0115	g14=g41
Test19	Wald	6.39	0.0115	g15=g51
Test20	Wald	6.39	0.0115	g16=g61
Test21	Wald	5.27	0.0217	g23=g32
Test22	Wald	3.40	0.0652	g24=g42
Test23	Wald	6.39	0.0115	g25=g52
Test24	Wald	6.39	0.0115	g26=g62
Test25	Wald	6.39	0.0115	g34=g43
Test26	Wald	6.39	0.0115	g35=g53
Test27	Wald	0.14	0.7108	g36=g63
Test28	Wald	6.39	0.0115	g45=g54
Test29	Wald	6.39	0.0115	g46=g64
Test30	Wald	6.39	0.0115	g56=g65

Table 44 Expenditure Elasticities and Uncompensated Own-Price and Cross-Price Demand Elasticities Estimated through Log Hedonic-Barten Synthetic Model (with Exponential Closeness Index)

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk	expenditure
almond milk	-0.1207**	-0.2065***	-0.0803***	-1.8828***	-0.8457***	-1.2532***	-0.9650***	5.0273***
	0.0519	0.0435	0.0289	0.4043	0.1766	0.2686	0.2022	1.0813
soy milk	-0.0429***	-0.4178***	-0.0251***	-3.3782***	-1.4706***	-2.2441***	-1.6854***	9.1810***
	0.0061	0.0418	0.0048	0.4514	0.1960	0.2998	0.2246	1.2069
rice milk	-0.1736***	-0.2544***	0.0206	-1.2169	-0.5813	-0.8296	-0.6738	3.0241
	0.0635	0.0839	0.1215	0.9092	0.3922	0.6026	0.4491	2.4627
2% milk	-0.0028***	-0.0160***	-0.0016***	-0.4166***	-0.1026***	-0.1566***	-0.1175***	0.8015***
	0.0004	0.0019	0.0003	0.0278	0.0120	0.0184	0.0138	0.0714
1% milk	-0.0046***	-0.0257***	-0.0026***	-0.3765***	-0.3318***	-0.2502***	-0.1902***	1.1771***
	0.0005	0.0021	0.0004	0.0320	0.0230	0.0213	0.0159	0.0864
fat-free milk	-0.0023***	-0.0122***	-0.0015***	-0.1767***	-0.0770***	-0.2873***	-0.0882***	0.6436***
	0.0004	0.0018	0.0003	0.0273	0.0119	0.0213	0.0136	0.0687
whole milk	-0.0024***	-0.0105***	-0.0017***	-0.1445***	-0.0652***	-0.0961***	-0.2476***	0.5565***
	0.0006	0.0024	0.0005	0.0362	0.0158	0.0240	0.0242	0.0949

Table 45 Compensated Own-Price and Cross-Price Demand Elasticities Estimated through Log Hedonic-Barten Synthetic Model (with Exponential Closeness Index)

	almond milk	soy milk	rice milk	2% milk	1% milk	fat-free milk	whole milk
almond milk	-0.1036**	-0.2567***	-0.0727**	-0.0033	-0.0294	-0.0048	-0.0294
	0.0510	0.0397	0.0289	0.0278	0.0229	0.0195	0.0246
soy milk	-0.0117**	-0.1936***	-0.0112**	0.0543***	0.0203***	0.0357***	0.0232***
	0.0049	0.0258	0.0045	0.0084	0.0044	0.0056	0.0051
rice milk	-0.1633**	-0.1805**	0.0251	-0.0863	-0.0902*	-0.0787*	-0.1110*
	0.0648	0.0731	0.1215	0.0601	0.0469	0.0483	0.0568
2% milk	0.0000	0.0035***	-0.0004	-0.1169***	0.0276***	0.0425***	0.0317***
	0.0003	0.0005	0.0002	0.0128	0.0032	0.0048	0.0036
1% milk	-0.0006	0.0031***	-0.0008*	0.0635***	-0.1407***	0.0421***	0.0289***
	0.0005	0.0007	0.0004	0.0073	0.0167	0.0049	0.0041
fat-free milk	-0.0001	0.0035***	-0.0005*	0.0639***	0.0275***	-0.1274***	0.0316***
	0.0003	0.0006	0.0003	0.0073	0.0032	0.0150	0.0036
whole milk	-0.0005	0.0030***	-0.0009*	0.0636**	0.0252***	0.0421***	-0.1441***
	0.000419	0.000625	0.000431	0.00704	0.00333	0.00468	0.0152

CHAPTER V

CONCLUSIONS

In terms of rate of growth, dairy alternative beverage market in the United States has surpassed the growth of conventional milk market in the recent years. The ongoing competition between dairy and dairy alternatives is expected to intensify over the next several years as consumers grow more comfortable with milk alternatives and criticism of dairy foods continue to grow. Sales of plant-based dairy alternative beverages, especially almond milk, show no signs of slowing and new alternative sources, such as rice milk and coconut milk, are expected to drive the alternative segment even faster and higher over the next several years.

The main objective of the work is to evaluate the consumer demand for dairy milk and milk alternatives using cutting-edge modifications to demand system models. We take product characteristics approach to study consumer demand for seven milk products. Income and price elasticities were estimated according to the economic activity during the period 2004-2015.

In this work, we have introduced qualitative factors through Hedonic Metrics method (HM) in approximating the elasticities estimated by the Barten synthetic model. Our method is based on two-stage estimation where we first allocated each differentiated product into a unique hedonic quality space. Next we estimate demand elasticities using distances between each product.

The Hedonic Metric estimations are based on qualitative distances calculated as the sum of the pair wise differences in the value added of each attribute for each product type. Therefore, it eliminates the need to search for significant characteristics which is required in the hedonic distance model.

Both linear and semi-log hedonic pricing models fit the data of seven products very well. Based on the estimates from hedonic pricing models, we constructed hedonic distance matrixes from which we derive the parameters used to reparamerize Barten synthetic model. The estimated parameters from Barten synthetic model are greatly reduced and are significant. Besides, the homogeneity and symmetry test results are as expected. Soy milk has the highest own-price elasticity. Inelastic demand of all three types of dairy alternative beverages means that they are not very sensitive to price changes. Soy milk is found to be substitute for all four types of conventional milk products and vise-versa. This provides an explanation of the consumption trend toward plant-based beverages. Three dairy alternative beverages are complements between each other.

Hedonic distance matrix approach is practical, eliminates the need to search for significant characteristics and has a stronger theoretical foundation, thus alleviating the ambiguity, while significantly reducing the number of parameters to estimate. Also, we applied a new method to acquire the own closeness index of each product. The algorithm is not based on the inverse of Euclidean distance, because it turns out to be all ones, which causes the singularity problem and thus biased estimates when we estimate the elasticities through the demand system.

The metric model applied in this paper can be applied to any market where product differentiation exists. Although we applied our model to dairy milk and milk alternative beverages at retail level, our model makes it possible to estimate the elasticities between differentiated products in any market. In many industries such as the electronic appliance industry or other agricultural food industry, both close and distant competition can be observed in multi-dimensional quality space, which is difficult to model. By implementing the qualitative factors within our approach, we can accommodate this behavior in a robust and simple way.

Consumer demand estimation results solidify the substitution effect between dairy alternative beverages and dairy milk product, which offers valid explanations of the trend or phenomenon of shifting consumption from dairy milk to dairy alternative beverages.

CHAPTER VI

LIMITATIONS AND FUTURE WORK

Limitations

Due to data limitations, we can only use pooled UPC level information to estimate the hedonic pricing models. Because the dairy alternative beverages are starting to gain ground in the recent ten years, adequate purchase observations were not available at the beginning of the time period pertaining to this study. Another reason is that, we need variations on the nutritional attributes, however household level data cannot guarantee enough variability. Therefore, we consider pooled data which can not only capture variability of the nutritional attributes but also enable us to expand the time period to be considered from year 2004 to 2015.Besides, the information about nutritional data is very scarce and limited for dairy alternative beverages in the Nielsen HomeScann database and also in the USDA nutritional database. Therefore, bulk of the nutritional data was collected from product labels. Besides, data limitations have constrained our selection of related dairy alternative beverages from which we can only include soy milk, almond milk and rice milk.

Secondly, compared to conventional milk products, dairy alternative beverages are becoming prominent in recent years, especially rice milk, so the market shares of these beverages could be quite small. The small market shares and high prices of these beverages might result in higher expenditure elasticities. Therefore, the estimated results of demand could possibly be more definitive and convincing if data after the year 2010 were used.

Thirdly, because the dataset is constructed based on UPC level, we cannot capture households' demographic information in estimating hedonic pricing models.

Future Work

Although almond milk and soy milk are the fastest growing categories, there exist numerous other products such as coconut milk, rice milk, cashew nut milk, hazelnut milk, etc. According to Nielsen, almond milk is now the America's most favorite milk substitute, and its estimated sales growth reaches up to 250 per cent over the five years' period from 2010 to 2015. Economic theory and industrial experience all suggest that the market structure of an industry strongly influences the competitive behavior of its member firms and the performance outcomes- prices, profits, output, etc. in its market. All other things being equal, prices are higher and price-cost margins (PCM) are wider under the conditions of monopoly than under the conditions of competition. The dairy alternative beverage market is characterized as high PCM and wider price margins with several well-known manufacturers and their branded products and a group of private label brands. They compete strategically for market share by differentiating their products by brand, price, advertising, promotion, positioning and merchandising. All the characteristics above have made this market a classic example of a monopolistically competitive market where rivalry is channeled into advertising and new product introduction and price competition is approximately cooperative.

While empirical studies of market power in agricultural markets are common, relatively few research concentrates on issues related to the market structure and market power of dairy milk market (see for example Moore and Clodius, 1962; Masson and Eisenstat (1980); Madhavan, Masson, and Lesser, 1994; Cakir and Balagtas, 2012; Shields, 2010). Due to the complicated market structure and the absence of research in the strategy conducted by dairy alternative firms, Nevo (2000) type of model can be developed to recover Price-cost margins of dairy alternative products without observing actual cost.

Another fruitful area of research is to use the demand estimates from this study in a merger simulation type of study involving several manufacturers in the dairy alternative beverage industry.

REFERENCES

Alston, J.M., Foster, K.A. and Gree, R.D., 1994. Estimating elasticities with the linear approximate almost ideal demand system: some Monte Carlo results. *The review of Economics and Statistics*, pp.351-356.

Alston, J.M., Sumner, D.A. and Vosti, S.A., 2008. Farm subsidies and obesity in the United States: National evidence and international comparisons. *Food Policy*, *33*(6), pp.470-479.

Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. *Econometrica: Journal of the Econometric Society*, pp.997-1016.

Amemiya, T., 1974. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica: Journal of the Econometric Society*, pp.999-1012.

Anderson, S.P., De Palma, A. and Thisse, J.F., 1989. Demand for differentiated products, discrete choice models, and the characteristics approach. *The Review of Economic Studies*, *56*(1), pp.21-35.

Anglin, P.M. and Gencay, R., 1996. Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, pp.633-648.

Ashenfelter, O. and Hosken, D., 2008. *The effect of mergers on consumer prices: Evidence from five selected case studies* (No. w13859). National Bureau of Economic Research.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John van Reenen. 2017. "Concentrating on the Fall of the Labor Share." *American Economic Review Papers and Proceedings*, vol. 107, pp. 180–185.

Baker, J.B., 2017. Market power in the US economy today. *Washington Center for Equitable Growth, available at http://equitablegrowth. org/research-analysis/marketpower-in-the-us-economy-today*.

Baker, J.B. and Bresnahan, T.F., 1988. Estimating the residual demand curve facing a single firm. *International Journal of Industrial Organization*, *6*(3), pp.283-300.

Baker, J.B. and Bresnahan, T.F., 1985. The gains from merger or collusion in product-differentiated industries. *The Journal of Industrial Economics*, pp.427-444.

Barten, A.P., 1993. Consumer allocation models: choice of functional form. *Empirical Economics*, *18*(1), pp.129-158.

Bartik, T.J., 1987. The estimation of demand parameters in hedonic price models. *Journal of political Economy*, *95*(1), pp.81-88.

Berry, S., Levinsohn, J. and Pakes, A., 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pp.841-890.

Berry, S.T., 1994. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pp.242-262.

Bloomberg; Nielsen, 2016. Sales growth of milk in the United States in 2016, by category [Online] Available: https://www.statista.com/statistics/548364/us-sales-growth-of-milk-by-category/.

Bontemps, C., Simioni, M. and Surry, Y., 2008. Semiparametric hedonic price models: assessing the effects of agricultural nonpoint source pollution. *Journal of Applied Econometrics*, *23*(6), pp.825-842.

Borgatti, S.P., 1995. Centrality and AIDS. *Connections*, *18*(1), pp.112-114.

Breustedt, G., Latacz-Lohmann, U. and Tiedemann, T., 2011. Organic or conventional? Optimal dairy farming technology under the EU milk quota system and organic subsidies. *Food Policy*, *36*(2), pp.223-229.

Bryant, H.L., Gogichaishvili, I., Anderson, D., Richardson, J.W., Sawyer, J., Wickersham, T. and Drewery, M.L., 2012. The value of post-extracted algae residue. *Algal Research*, *1*(2), pp.185-193.

Cakir, M. and Balagtas, J.V., 2012. Estimating market power of US dairy cooperatives in the fluid milk market. *American Journal of Agricultural Economics*, *94*(3), pp.647-658.

Capps Jr, O., Church, J. and Love, H.A., 2003. Specification issues and confidence intervals in unilateral price effects analysis. *Journal of Econometrics*, *113*(1), pp.3-31.

Carbonneau, P. and Meilhac, L., 2016. Got growth? Opportunities and challenges for US Dairy industry. *McKinsey&Company, Ed*, pp.1-15.

Choi, H.J., Wohlgenant, M.K. and Zheng, X., 2013. Household-level welfare effects of organic milk introduction. *American Journal of Agricultural Economics*, *95*(4), pp.1009-1028.

Clapp, J.M., 2004. A semiparametric method for estimating local house price indices. *Real Estate Economics*, *32*(1), pp.127-160.

Coate, M.B. and Kleit, A. eds., 1996. *The Economics of the Antitrust Process* (Vol. 22). Springer Science & Business Media.

Combris, P., Lecocq, S. and Visser, M., 1997. Estimation of a hedonic price equation for Bordeaux wine: does quality matter?. *The Economic Journal*, *107*(441), pp.390-402.

Cooper, C.H., 2015. Spatial localization of closeness and betweenness measures: a self-contradictory but useful form of network analysis. *International Journal of Geographical Information Science*, *29*(8), pp.1293-1309.

Copeland, A. and Dharmasena, S., 2016. Impact of increasing demand for dairy alternative beverages on dairy farmer welfare in the United States.

Cotterill, R.W., Rabinowitz, A.N. and Tian, L., 2003. Milk Market Channel Structure: Its Impact on Farmers and Consumers, and the Inadequacies of Antitrust Enforcement as a Foundation for Dairy Policies: Evidence from the Northeast Dairy Industry. *Testimony before the United States Senate Committee on the Judiciary, Washington DC, October*, *23*.

Crooke, P., Froeb, L., Tschantz, S. and Werden, G.J., 1999. Effects of assumed demand form on simulated postmerger equilibria. *Review of Industrial Organization*, *15*(3), pp.205-217.

Cropp, R.A., 2002. Dairy Cooperatives: Structural Change and Operations. *University of Wisconsin Center for Cooperatives: University of Wisconsin—Extension Bulletin*, (4).

Cropper, M.L., Deck, L.B. and McConnell, K.E., 1988. On the choice of funtional form for hedonic price functions. *The Review of Economics and Statistics*, pp.668-675.

Cotterill, R.W., Rabinowitz, A.N. and Tian, L., 2003. Milk Market Channel Structure: Its Impact on Farmers and Consumers, and the Inadequacies of Antitrust Enforcement as a Foundation for Dairy Policies: Evidence from the Northeast Dairy Industry. *Testimony before the United States Senate Committee on the Judiciary, Washington DC, October*, *23*.

Dhar, T. and Foltz, J.D., 2004. Is soy milk? The economics of the soy milk market. In *American Association of Agricultural Economics Meeting, Denver, USA*.

Dharmasena, S. and Capps, O., 2014. Unraveling demand for dairy-alternative beverages in the United States: The case of soymilk. *Agricultural and Resource Economics Review*, *43*(1), pp.140-157.

Dharmasena, S., Davis, G.C. and Capps Jr, O., 2014. Partial versus General Equilibrium Calorie and Revenue Effects Associated with a Sugar-Sweetened Beverage Tax. *Journal of Agricultural & Resource Economics*, *39*(2).

Dohnal, V., 2004. Indexing structures for searching in metric spaces. *Disertační práce, Fakulta informatiky, MU Brno*.

Dong, D., Gould, B.W. and Kaiser, H.M., 2004. Food demand in Mexico: an application of the Amemiya-Tobin approach to the estimation of a censored food system. *American Journal of Agricultural Economics*, *86*(4), pp.1094-1107.

Dubé, J.P., 2005. Product differentiation and mergers in the carbonated soft drink industry. *Journal of Economics & Management Strategy*, *14*(4), pp.879-904.

Cakir, M. and Balagtas, J.V., 2012. Estimating market power of US dairy cooperatives in the fluid milk market. *American Journal of Agricultural Economics*, *94*(3), pp.647-658.

Cardell, S., 1989. “The Hedonic Demand Model and Some Other Extensions of Multinomial Logit.” Ph.D. dissertation, Harvard University.

Chamberlain, G., 1982. Multivariate regression models for panel data. *Journal of* *econometrics*, *18*(1), pp.5-46.

Cooper, C.H., 2015. Spatial localization of closeness and betweenness measures: a self-contradictory but useful form of network analysis. *International Journal of Geographical Information Science*, *29*(8), pp.1293-1309.

Epple, D., 1987. Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products. *Journal of political economy*, *95*(1), pp.59-80.

Franck, C., Grandi, S.M. and Eisenberg, M.J., 2013. Agricultural subsidies and the American obesity epidemic. *American journal of preventive medicine*, *45*(3), pp.327-333.

G. Sabidussi, “The centrality index of a graph,” *Psychometrika, vol. 31, no. 4,* pp. 581–603, 1966

Gasmi, F., Laffont, J.J. and Vuong, Q., 1992. Econometric Analysisof Collusive Behaviorin a Soft-Drink Market. *Journal of Economics & Management Strategy*, *1*(2), pp.277-311.

Gasmi, F. and Vuong, Q.H., 1988. *An econometric analysis of some duopolistic games in prices and advertising* (No. m8903).

Gould, B.W., 1996. Factors affecting US demand for reduced-fat fluid milk. *Journal of Agricultural and Resource Economics*, pp.68-81.

Gould, B.W., 2010. Consolidation and concentration in the US dairy industry. Choices, 25(2), pp.1-15.

Griliches, Z., 1961. Hedonic price indexes for automobiles: An econometric of quality change. In The Price Statistics of the Federal Goverment (pp. 173-196). NBER.

Grand View Research; Statista estimates, 2016. Market value of dairy milk alternatives worldwide from 2013 to 2024, by category (in million U.S. dollars) [Online] Available: https://www.statista.com/statistics/693015/dairy-alternatives-global-sales-value-by-category/.

Grand View Research; Statista; Statista estimate, 2016. Sales value of rice milk in the United States from 2013 to 2024 (in million U.S. dollars) [Online] Available: https://www.statista.com/statistics/693121/us-rice-milk-sales-value/.

Gul, A., Akbay, C. and Direk, M., 2006. Sunn pest control policies and effect of sunn pest damage on wheat quality and price in Turkey. *Quality and Quantity*, *40*(3), pp.469-480.

Guha, S., Rastogi, R. and Shim, K., 1999, March. ROCK: A robust clustering algorithm for categorical attributes. In Data Engineering, *1999. Proceedings., 15th International Conference on* (pp. 512-521). IEEE.

Gulseven, O. and Wohlgenant, M., 2015. A quality-based approach to estimating quantitative elasticities for differentiated products: an application to retail milk demand. *Quality & Quantity*, *49*(5), pp.2077-2096.

Gustafson, C.R., Lybbert, T.J. and Sumner, D.A., 2016. Consumer sorting and hedonic valuation of wine attributes: exploiting data from a field experiment. *Agricultural economics*, *47*(1), pp.91-103.

Halvorsen, R. and Pollakowski, H.O., 1981. Choice of functional form for hedonic price equations. *Journal of urban economics*, *10*(1), pp.37-49.

Hanemann, W.M., 1984. Discrete/continuous models of consumer demand. *Econometrica: Journal of the Econometric Society*, pp.541-561.

Hausman, J.A. and Leonard, G.K., 1996. Economic analysis of differentiated products mergers using real world data. *Geo. Mason L. Rev.*, *5*, p.321.

Hausman, J., Leonard, G. and Zona, J.D., 1994. Competitive analysis with differenciated products. *Annales d'Economie et de Statistique*, pp.159-180.

Hendel, I., 1999. Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies*, *66*(2), pp.423-446.

Hendrickson, M., Heffernan, W.D., Howard, P.H. and Heffernan, J.B., 2001. Consolidation in food retailing and dairy. *British Food Journal*, *103*(10), pp.715-728.

Johnston, J.W., 1976. *Similarity indices I: What do they measure* (No. BNWL--2152 (ADD. 1)). Battelle Pacific Northwest Labs..

Kim, E.H. and Singal, V., 1993. Mergers and market power: Evidence from the airline industry. *The American Economic Review*, pp.549-569.

Kinoshita, J., Suzuki, N., Kawamura, T., Watanabe, Y. and Kaiser, H.M., 2001. Estimating own and cross brand price elasticities, and price–cost margin ratios using store-level daily scanner data. *Agribusiness*, *17*(4), pp.515-525.

Ladd, G.W. and Martin, M.B., 1976. Prices and demands for input characteristics. *American Journal of Agricultural Economics*, 58(1), pp.21-30.

Ladd, G.W. and Suvannunt, V., 1976. A model of consumer goods characteristics. *American Journal of Agricultural Economics*, *58*(3), pp.504-510.

Lancaster, K.J., 1966. A new approach to consumer theory. *Journal of political economy*, *74*(2), pp.132-157.

Lee, L.F. and Pitt, M.M., 1986. Microeconometric demand system with binding nonnegativity constraints: the dual approach. *Econometrica: Journal of the Econometric Society*, pp.1237-1242.

Lien, G. and Hardaker, J.B., 2001. Whole-farm planning under uncertainty: impacts of subsidy scheme and utility function on portfolio choice in Norwegian agriculture. *European review of agricultural economics*, *28*(1), pp.17-36.

Li, J. and Dharmasena, S., 2016, May. Investigating Economic and Demographic Factors Affecting Consumer Demand for Coconut-milk in the United States. In *Selected paper presented at the Agricultural and Applied Economics Association annual meetings, Boston, Massachusetts*.

MacDonald, J.M., 2017. Consolidation, concentration, and competition in the food system. *Economic Review*, pp.85-105.

MarketsandMarkets (TM), 2017. Dairy Alternatives Market by Type (Soy, Almond, Coconut, Rice, Oat, Hemp), Formulation (Plain & Sweetened, Plain & Unsweetened, Flavored & Sweetened, Flavored & Unsweetened), Application (Food, Beverages), and Region - Global Forecast to 2022 [Online] Available: https:www.marketsandmarkets.com/Market-Reports/dairy-alternative-plant-milk-beverages-market-677.html.

McFadden, D., 1981. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, *198272*.

McFadden, D., 1978. Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior.

Madhavan, A.N., Masson, R.T. and Lesser, W.H., 1994. Cooperation for monopolization? An empirical analysis of cartelization. *The Review of Economics and Statistics*, pp.161-175.

Manchester, A.C. and Blayney, D.P., 1997. The Structure of Dairy Markets: Past, Present, Future. AER 757. Washington, DC: US Department of Agriculture. *Economic Research Service*.

Matsuda, T., 2005. Differential demand systems: A further look at Barten's synthesis. *Southern Economic Journal*, pp.607-619.

Masson, R.T. and Eisenstat, P.M., 1980. Welfare impacts of milk orders and the antitrust immunities for cooperatives. *American Journal of Agricultural Economics*, *62*(2), pp.270-278.

Meese, R. and Wallace, N., 1991. Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *Real Estate Economics*, *19*(3), pp.308-332.

Meyerhoefer, C.D., Ranney, C.K. and Sahn, D.E., 2005. Consistent estimation of censored demand systems using panel data. *American Journal of Agricultural Economics*, *87*(3), pp.660-672.

Miles, H., Schwager, S.J. and Lenz, J.E., 1995. Perceptual dimensions that influence consumers' choices of milk type for beverage use. *Agribusiness*, *11*(3), pp.263-272.

Moore, J.R. and Clodius, R.L., 1962. Market structure and competition in the dairy industry.

Nerlove, M., 1995. Hedonic price functions and the measurement of preferences: The case of Swedish wine consumers. *European economic review*, *39*(9), pp.1697-1716.

Nimon, W. and Beghin, J., 1999. Are eco-labels valuable? Evidence from the apparel industry.

American Journal of Agricultural Economics, *81*(4), pp.801-811.

Nielsen; WhiteWave Foods; Danone, 2017. Household penetration rate of plant-based beverages in the United States from 2010 to 2016 [Online] Available: https://www.statista.com/statistics/474461/us-household-penetration-rate-plant-based-beverages/.

Nordin, B.C., 2000. Calcium requirement is a sliding scale. *The American journal of clinical nutrition*, *71*(6), pp.1381-1383.

O. Ashenfelter & D. Hosken, Price Effects of Consumer Product Market Mergers (Princeton University, Working Paper, 2004).

Okamoto, K., Chen, W. and Li, X.Y., 2008, June. Ranking of closeness centrality for large-scale social networks. In *International Workshop on Frontiers in Algorithmics* (pp. 186-195). Springer, Berlin, Heidelberg.

Ollinger, M., Nguyen, S.V., Blayney, D., Chambers, B. and Nelson, K., 2005. *Structural change in the meat, poultry, dairy, and grain processing industries*. US Department of Agriculture, Economic Research Service.

Packaged Facts. "Dairy and Dairy Alternative Beverage Trends in the U.S. 4th Edition."

Palmquist, R.B., 1984. Estimating the Demand for the Characteristics of Housing. *The Review of Economics and Statistics*, pp.394-404.

Parmeter, C.F., Henderson, D.J. and Kumbhakar, S.C., 2007. Nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, *22*(3), pp.695-699.

Perali, F. and Chavas, J.P., 2000. Estimation of censored demand equations from large cross-section data. *American Journal of Agricultural Economics*, *82*(4), pp.1022-1037.

Perloff, J.M., Karp, L.S. and Golan, A., 2007. *Estimating market power and strategies*. Cambridge University Press.

Peters, C., 2006. Evaluating the performance of merger simulation: Evidence from the US airline industry. *The Journal of law and economics*, *49*(2), pp.627-649.

Pinkse, J., Slade, M.E. and Brett, C., 2002. Spatial price competition: a semiparametric approach. *Econometrica*, *70*(3), pp.1111-1153.

Prager, R.A. and Hannan, T.H., 1998. Do substantial horizontal mergers generate significant price effects? Evidence from the banking industry. *The Journal of Industrial Economics*, *46*(4), pp.433-452.

Progressive Grocer; IRI, 2016. Refrigerated milk sales in the United States in 2016, by category (in million U.S. dollars) [Online] Available: https://www.statista.com/statistics/589326/us-sales-refrigerated-milk-by-category/.

Rasmussen, D.W. and Zuehlke, T.W., 1990. On the choice of functional form for hedonic price functions. *Applied Economics*, *22*(4), pp.431-438.

Rojas, C. and Peterson, E.B., 2008. Demand for differentiated products: Price and advertising evidence from the US beer market. *International Journal of Industrial Organization*, *26*(1), pp.288-307.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, *82*(1), pp.34-55.

Sexton, R.J. and Lavoie, N., 2001. Food processing and distribution: an industrial organization approach. *Handbook of agricultural economics*, *1*, pp.863-932.

Shields, D.A., 2010. Consolidation and concentration in the US dairy industry. *Congressional Research Service Report*, *41224*.

Shonkwiler, J.S. and Yen, S.T., 1999. Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics, 81(4)*, pp.972-982.

Silanikove, N., Leitner, G. and Merin, U., 2015. The interrelationships between lactose intolerance and the modern dairy industry: global perspectives in evolutional and historical backgrounds. *Nutrients, 7(9)*, pp.7312-7331.

Slade, M.E., 2006. *Merger Simulations of Unilateral Effects: What can we learn from the UK brewing Industry?*. Department of Economics, University of Warwick.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, pp.24-36.

Triplett, J.E., 1986. The economic interpretation of hedonic methods. *Survey of current business, 66(1)*, pp.36-40.

U.S. Government Accountability Office, *Concentration in Agriculture*, GAO-09-746R, June 30, 2009, p. 26, http://www.gao.gov/new.items/d09746r.pdf.

Van Dongen, S.M., 2000. *Graph clustering by flow simulation* (Doctoral dissertation).

Wales, T.J. and Woodland, A.D., 1983. Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics, 21*(3), pp.263-285.

Wang, W. and Tang, C.Y., 2014, June. Distributed computation of classic and exponential closeness on tree graphs. In *American Control Conference (ACC), 2014* (pp. 2090-2095). IEEE.).

Wang, Z., Scaglione, A. and Thomas, R.J., 2010, December. Electrical centrality measures for electric power grid vulnerability analysis. In *Decision and Control (CDC), 2010 49th IEEE Conference on* (pp. 5792-5797). IEEE.

Waugh, F.V., 1928. Quality factors influencing vegetable prices. *Journal of farm economics*, *10*(2), pp.185-196.

Weinberg, M.C. and Hosken, D., 2009. Using mergers to test a model of oligopoly. *Unpublished paper, Cornell Department of Economics*.

Weinberg, M., 2007. The price effects of horizontal mergers. *Journal of competition Law and Economics*, *4*(2), pp.433-447.

Werden, G.J., 1997. Demand Elasticites in Antitrust Analysis. *Antitrust LJ*, *66*, p.363.

Werden, G.J., 1997. Antitrust Analysis of Joint Ventures-An Overview. *Antitrust LJ*, *66*, p.701.

Wise, T.A., 2004. *The paradox of agricultural subsidies: measurement issues, agricultural dumping, and policy reform*. Tufts University.

Yen, S.T. and Lin, B.H., 2006. A sample selection approach to censored demand systems. *American Journal of Agricultural Economics*, 88(3), pp.742-749.

Yen, S.T., Lin, B.H. and Smallwood, D.M., 2003. Quasi-and simulated-likelihood approaches to censored demand systems: food consumption by food stamp recipients in the United States. *American Journal of Agricultural Economics*, *85*(2), pp.458-478.

Yen, S.T., Kan, K. and Su, S.J., 2002. Household demand for fats and oils: two-step estimation of a censored demand system. *Applied Economics*, *34*(14), pp.1799-1806.