VALUE OF PARSIMONIOUS NUTRITIONAL INFORMATION, CONSUMER-ORIENTED FOODS CLUSTER, AND PREDICTING FOOD PRICE

By

JISUNG JO

Bachelor of Science in Agricultural Economics
Pusan National University
Busan, Korea
2011

Master of Science in Agricultural Economics
Seoul National University
Seoul, Korea
2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2016

VALUE OF PARSIMONIOUS NUTRITIONAL VALUE

OF PARSIMONIOUS NUTRITIONAL INFORMATION,

CONSUMER-ORIENTED FOODS CLUSTER, AND

PREDICTING FOOD PRICE

Dissertation Approved:

Dr. Jayson L. Lusk

Dissertation Adviser

Dr. B. Wade Brorsen

Dr. Bailey Norwood

Dr. Lan Zhu

Name: Jisung Jo

Date of Degree: DECEMBER, 2016

Title of Study: VALUE OF PARSIMONIOUS NUTRITIONAL INFORMATION, PREDICTING FOOD PRICE, AND CONSUMER-ORIENTED FOODS CLUSTER

Major Field: AGRICULTURAL ECONOMICS

Abstract: This dissertation focuses on three topics that relate to consumer behavior and the food industry. The first chapter investigates consumers' beliefs about the tastiness and healthfulness of 173 food items in a framed field experiment. Using data collected from 129 food shoppers in Grenoble France, demand models are estimated to determine how choices change with the provision of objective health information. We elicit and convey health information using simple nutritional indices meant to lower search and cognitive processing costs. The results indicate that consumers are willing to pay for tastier foods and for healthier foods, particularly if the consumers have objective information on nutrient content. The estimates suggest that the value of the type of nutritional information provided in the experiment is €0.98 per day. The second chapter investigates USA, China, and Korea consumers' perceptions about the health, taste, and price of 60 different food items to determine country-specific food clusters before and after the provision of objective health information. Subsequent analysis relates cluster characteristics to purchase intentions. For Hedonic and Taste-oriented cluster products, Koreans' purchase intentions rise if the products are perceived as expensive before the provision of information; however the purchase intention of Americans and Chinese is not affected by beliefs about affordability. These results could help retailers in each country identify appropriate food groupings, from the consumers' perspective, to improve category management, marketing, and pricing. The last chapter explores whether unconventional consumer-oriented variables might be useful in predicting the Bureau of Labor Statistics (BLS) Food and Beverages Consumer Price Index (CPI). We determine the ability of an Internet search-based index related to food prices (the Google trends index) and a survey-based consumer sentiment index to predict changes in food-related BLS prices from January 2004 to July 2015. A vector autoregression (VAR) model has the best predictive performance with the moving window structure and a vector error correction model (VECM) performs best with the expanding window structure. Encompassing tests reveal that our model out-predicts USDA Economic Research Service food-related CPI forecasts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

VALUE OF PARSIMONIOUS NUTRITIONAL INFORMATION IN A FRAMED FIELD EXPERIMENT

Introduction

In the United States, nutrition labels on packaged foods have been mandatory for over 20 years. European countries have been slower to adopt mandatory labels, but various standards and voluntary programs exist. The laws in each country normally require some form of standardized nutrition labels. These labels provide a wealth of information about calories along with macro- and micro-nutrient content. In accordance with the prevalence of nutrient labeling use, there have been several studies on the effectiveness and value of nutrition labels (Drichoutis, Lazaridis, & Nayga, 2006; Drichoutis, Nayga, & Lazaridis, 2011; and Grunert & Wills, 2007). However, results of these studies differ by the types of food and nutrient information, and they often rely on self-reported label use. These studies have suggested, for example, that the provision of information has a positive effect on the consumption of healthy ingredients such as fiber and a negative effect on the consumption of less healthy ingredients like fat and cholesterol (Drichoutis et al., 2006). However, it might be possible that simplified label formats are even more effective, and in fact prior research has suggested that consumers prefer simplified front of pack information rather than complex nutrition labels (Gruner & Wills, 2007). This paper was designed to determine the effect of simple nutrient information on consumer choice in an experimental context involving real food and real money in a manner that allows us to estimate the economic value of nutritional information aggregated over an entire day's meal choices.

Typical label designs tend to rest on the assumption that more information is better and that consumers will rationally update their subjective beliefs in response to objective information provided. However, research in behavioral economics suggests that the way information is framed, subtle cues, prior beliefs, and the amount of information released can have substantive effects on consumer behavior (Kahneman & Tyersky, 2000; Rabin & Schrag, 1999; Wansink, 2004). In the context of food labels, this has led to public and private efforts to more succinctly convey nutritional information via traffic lights system (TLS) or front-of-package (FOP) labeling. Balcombe et al. (2010) found a strong preference on the part of consumers in the UK to reduce the quantity of any nutrient associated with a red light, indicating a food that is high in fat, sugar, or salt. Ellison, Lusk, and Davis (2014) showed that numeric labels did not influence food choice in a restaurant, but TLS caused restaurant patrons to select lower-calorie menu items. Also, Roberto et al. (2012) mentioned that listing calories per serving information on FOP labels can increase knowledge and influence purchasing behavior. In fact, the US Food and Drug Administration (FDA) recently redesigned mandatory nutrition labels to more prominently emphasize overall calorie content and added sugars (Food and Drug Administration, 2014).

These previous papers suggest simple nutrient labeling is likely preferable to complex information. These findings prompted us to explore a simple form of nutrient information conveyed by two nutritional indices. One index provides information on the content of beneficial nutrients and the other provides information on less healthy nutrients; these simplified indices represent a succinct way to convey complex nutrient information (which previous research suggests reduces effectiveness) in a manner that is perhaps more transparent than TLS. Moreover, the index approach can be broadly and consistently applied across a wide array of foodstuffs.

Many of the previous studies on the effects of nutritional labeling tend to use consumers' self-reports of label use in surveys (Kreuter et al., 1997; Garretson & Burton, 2000; Derby & Levy, 2001). Unfortunately, such self-reports can be unreliable and may be endogenously determined with other factors, such as health consciousness and nutritional knowledge. To address some of these concerns, some research has studied consumers' actual purchases in a retail setting before and after the provision of nutritional information (Teisl, Bockstael, & Levy, 2001). Such studies are typically limited to a handful of product categories, and as such, do not provide a comprehensive measure of the value of information to a shopper. Moreover, such studies often lack data on consumers' prior nutritional beliefs and may attribute changes in choice solely to nutrition, when in fact nutritional labels and claims may change taste perceptions (Kiesel & Villas-Boas, 2013).

Rather than relying on self-reports of label use, as has often been the case with prior research (Drichoutis et al., 2005; Derby & Levy, 2001; Feunekes et al., 2008; Gracia et al., 2007), we conduct a framed field experiment in which consumers make non-hypothetical food choices before and after the provision of information. Unlike prior research based on actual consumer purchases (e.g., Weaver & Finke, 2003), our experimental setting enables us to measure consumers' prior beliefs about the tastiness and nutritional content of foods. This allows us to better understand how consumers update their perceptions of the healthiness of food and how they sometimes tradeoff health for taste (Drichoutis et al., 2006; Smith, 2004). Akin to Teisl, Bockstael, and Levy (2001), we provide an explicit estimate of the economic value of the nutritional information conveyed in the indices, but unlike their analysis, our experimental approach allows us to estimate this value over a very wide range of food products, which allows us to arrive at an aggregate value of information irrespective of the particular types of foods chosen by a particular consumer.

The experiment was not conducted in a grocery store; however, by moving to a more controlled (though still non-hypothetical-real food-real money) environment, we are able to more conclusively identify the effects of interest. That is, our field experiment attempts to mimic a real market situation and has many advantages. First, we observe respondents' choice behaviors directly in treatment and control situations where we can be sure confounding factors did not enter. Second, although 173 food items used in our experiment represent a small portion of the options in the real world sold by grocery stores, the number of food options reasonably reflect the categories of choices available to respondents in the grocery store without providing overwhelming differentiation (e.g., apple cinnamon cheerios, honey nut cheerios medley crunch, chocolate cheerios, and multi grain peanut butter cheerios). This allows us to focus on cross-category substitution rather than within-category substitution. The 173 food items were chosen on the basis of average consumption by French people and in consultation with prominent nutritionists. Lastly, the repeated food choices under different labels and prices is not unlike what occurs in actual market situations. People usually shop for food repeatedly, and are confronted with food price changes in the real world. Moreover, Chang et al. (2009) has found non-hypothetical laboratory experiments have high external validity, leading to accurate prediction of grocery store market shares. Nonetheless, we suggest the resulting value of information we obtain is likely to represent an upper-bound measure because our within-subject, controlled environment is likely to focus more attention on the labels than might be the case in a "noisier" field environment.

Our research additionally builds on previous studies in other important ways. Teisl et al. (2001) showed that although nutrient labeling affected purchase behavior (and thus has positive value), it did not necessarily increase consumption of healthy food. This is because provision of health information can also signal information about taste. If people tend to associate more tasty food with less healthy food, the provision of health information could have unintended effects (Tepper & Trail, 1998; Raghunathan, Naylor, & Hoyer, 2006; Mai & Hoffmann, 2014). In accordance with this previous research, by asking consumers to rate the taste of each of the 173 food items on a -5 to +5 scale, where -5 represents distasteful and +5 represents delicious, our study includes taste as a utility driver. This allows us to study the impact of health information to deal with psychological effects when people face the health-related information.

In the following section, we describe our experiment. The economic approach used to estimate demand is then described. Results are then discussed, and the last section concludes the discussion of this study.

Experiment

The data for this study comes from a framed field experiment conducted in Grenoble, France. One hundred and twenty nine women between the ages of 18 and 76 participated in the study. We recruited only women because they are the primary food shoppers in most French households. Subjects were recruited by placing announcements around town; subjects were offered a 20€ show-up fee for participation. During the introductory phase, the experimenter made sure the participants understood this amount of money (20€) was unrelated with the following tasks of the experimental session.

The experiment requested the participants to choose all the foods and drinks they desire to purchase for breakfast, lunch, and dinner for a given day using a hand-held scanner and a computer interface. The choices were repeated under three treatments or "days" (Figure 1-1 summarizes the steps in the experiment).[1] We utilize a within-subject design so that each subject makes a day's worth of food choices in three different treatments. In each treatment, subjects were given a catalog from which they could select from among 173 different food items, each shown with a photo and corresponding price, using a handheld scanner. For anonymity, an identification number was the only way the participants could be identified in the experiment.

During the food choice task, participants were not restrained in their spending. Neither upper limits nor lower limits were set. This is important for three main reasons. First, we did not want to omit income effects. With a fixed budget constraint, only substitution effects would have been observable. Second, forcing consumers to fully spend a fixed endowment can induce a variety of incentives that are antithetical to truthful preference revelation (Fischer, 2014). Lastly, we wanted, as much as possible, to avoid endowment effect generated by the initial compensation. With no budget restriction, the money saved in the lab can be spent outside the lab and the money spent in the lab is lost outside the lab. By doing so, we could maintain opportunity cost and experimental money as truly real money.

Prior to making food choices, respondents were asked to rate each food's taste on a scale ranging from -5 to +5, where -5 represents distasteful and +5 represents delicious. After indicating the taste perceptions of each of the 173 food items, the participants began treatment 1

1 We did not randomize the order. However, no information whatsoever has been given during this task. Therefore, participants could not learn from their previous decisions. The only learning process possible is some kind of learning-by-doing, but it is difficult to imagine how such repetition could improve knowledge without any feedback between decisions.

(or “day 1”) in which they picked which items (and how much) they wanted to satisfy a day’s worth of food consumption.

The initial “day 1” food choices were based on the individuals’ subjective (and implicit) health beliefs. Between days 1 and 2, we sought to measure those subjective health beliefs and also to provide objective information about each of the 173 foods. The beliefs were measured by asking respondents to pick the quadrant in the SAIN (Nutrient Adequacy Score for Individual foods) and LIM (for Limited Nutrient) table (Figure 1-2) that best described where they thought each food fit. The SAIN and LIM are nutrient profiling models and indices introduced by the French Food Safety Agency. The SAIN score is a measure of “good” nutrients calculated as an un-weighted arithmetic mean of the percentage adequacy for five positive nutrients: protein, fiber, ascorbic acid, calcium, and iron. The LIM score is a measure of “bad” nutrients calculated as the mean percentage of the maximum recommended values for three nutrients: sodium, added sugar, and saturated fatty acid.[2] Since indices help reduce search costs, displaying the information in the form of an index is a way to make the information available in an objective way but also allows consumers to better compare the many alternative products in their choice set.

---

2 The SAIN score is calculated as

$$SAIN_i = \frac{\left(\frac{Protein_i}{65} + \frac{Fiber_i}{25} + \frac{Ascorbic\ acid_i}{0.11} + \frac{Calcium_i}{0.9} + \frac{Iron_i}{0.0125}\right) \times \frac{100}{E_i}}{5} \times 100$$

where *Protein*, *Fiber, Ascorbic acid, Calcium*, and *Iron* are the quantities (g, mg or µg) of each nutrient in 100g of food *i*, *E* is the energy content of 100g of food *i* (kcal/100g), and 65,25,0.11,0.9, and 0.0125 are the daily recommended values (g) for each nutrient, respectively.

The LIM score is calculated as

$$LIM_i = \frac{\left(\frac{Saturated\ Fatty\ Acid_i}{22} + \frac{Added\ Sugar_i}{50} + \frac{Sodium_i}{3.153}\right)}{3} \times 100$$

where *Saturated fatty acid, Added Sugar,* and *Sodium* are the quantities (g and mg) of each nutrient in 100g of food *i*, and 22, 50, and 3.153 are the daily maximal recommended values (g) for each nutrient.

Figure 1-2 shows that each food can be placed in one of four quadrants depending on whether the food is high or low in the SAIN and LIM indices. Darmon et al. (2009) determined the “high” and “low” acceptability thresholds for SAIN and LIM as 5 and 7.5, respectively. Food in quadrant 2, where SAIN is high and LIM is low, is considered healthy food. Most fruits and vegetables are included in quadrant 2. Quadrant 4 has a low SAIN and high LIM score, which means foods in this quadrant are unhealthy; the category includes foods such as snacks, cakes, and sweets. Food in quadrant 1 is nutritionally beneficial, but should be eaten occasionally and in small quantities. Ham, red meats, and some cheeses are in quadrant 1. Lastly, bread, pasta, and rice are included in neutral quadrant 3, which denotes a low SAIN and low LIM score. Though these products can be consumed regularly because of their low nutrient intake, they must be accompanied with high nutrient food.

Respondents were incentivized to carefully answer the perceived healthiness of each food. In particular, they were given 0.05€ for each food they placed in the correct quadrant (thus, each participant could earn up to 173*0.05=8.65€ if they correctly placed each food item in the proper category). Immediately after indicating the health quadrant for a particular food, the software program indicated whether the answer was correct or incorrect. If the answer was incorrect, then the respondent was informed as to which quadrant the food actually belonged. This process was completed for all 173 foods so that for each food we have the individuals' implicit subjective belief, and we are also able to easily convey objective health information for all foods.

After completing all the health ratings (and receiving information on the healthiness) for each food, subjects moved to treatment 2. In treatment 2 (or “day 2”), subjects repeated their purchases. The task was the same as in treatment 1, except in this case the individuals had objective information of where each of the 173 foods fit in the SAIN/LIM matrix in Figure 1-2.

The final, third treatment was the same as treatment 2 except the prices of healthy foods, according to the SAIN/LIM indices were reduced, and the prices of the unhealthy foods according to the SAIN/LIM indices were increased.[3] Thus, the data set consists of choices among 173 foods in three treatments that varied by the provision of nutrition information and price.

To incentivize the choices, one of the three days was randomly selected as binding. Then, for the binding day, around 50 food items were selected as binding, and if a participant selected one of these binding food items in the binding day, they purchased it at the stated price. Because participants did not know which food day or which food items would ultimately be binding, they had an incentive to carefully consider each choice and respond in a manner that accurately reflected their true preferences.

Econometric Methods

Data are pooled from treatments (or days) 1, 2, and 3 to estimate an attribute-based, random utility model (RUM) of McFadden (1973). The systematic utility consumer *i* derives from product *k* in treatment *t* is

$$(1)\quad V_{ikt} = \beta_1 Cereal_k + \beta_2 Dairy_k + \beta_3 Fruit_k + \beta_4 Meat_k + \beta_5 Mixed_k + \beta_6 Snack_k$$

3 Meuller et al. (2016) used a similar experiment set up, and they changed food prices to study the effects of unhealthy food taxes and healthy food subsidies, and we followed their approach. For the purposes of the present inquiry, we simply need some price variation so we can clearly identify the price coefficient in the econometric model, and our design allows us to do that.

$$+\beta_7 Veggie_k + \beta_8 Taste_{ik} + \beta_9 Healthy\_before_{ik} + \beta_{10} Unhealthy\_before_{ik}$$

$$+ \beta_{11} Healthy\_after_{ik} + \beta_{12} Unhealthy\_after_{ik} + \beta_{13} Price_{kt},$$

where $Cereal_k$, $Dairy_k$, $Fruit_k$, $Meat_k$, $Mixed_k$, $Snack_k$ and $Veggie_k$ are the binary variables indicating food *k*'s type, where *k*=1,2,…,173; $Taste_{ik}$ is the i^{th} individual's perceived taste of the k^{th} food item where *i*=1,2,…,129; $Healthy\_before_{ik}$ is a dummy variable describing whether the i^{th} individual perceives that food *k* is healthy in treatment 1; $Unhealthy\_before_{ik}$ is a dummy variable describing whether the i^{th} individual perceives food *k* to be an unhealthy food in treatment 1; $Healthy\_after_{ik}$ is a dummy variable denoting whether food *k* is truly a healthy food (in treatments 2 and 3 after information); $Unhealthy\_after_{ik}$ is a dummy variable indicating whether food is truly an unhealthy food (in treatments 2 and 3); $Price_{kt}$ is the price of the k^{th} food item in treatment *t* where *t*=1,2,3; and $\beta_1, \ldots, \beta_{13}$ are the coefficients (marginal utilities) for each explanatory variable.[4]

We categorized the healthiness of a food based on where it fell on the nutrient indices as shown in Figure 1-1. The dummy variables $Healthy\_before_{ik}$ and $Unhealthy\_before_{ik}$ represent whether, in treatment 1, subjects believed a food was from quadrant 2 or quadrant 4, respectively. Also, the food items from quadrant 1 and quadrant 3 are considered $Mid-level\ Healthy\_before_{ik}$. In treatments 2 and 3, subjects have access to objective information on each food's placement in the SAIN/LIM matrix. The variables $Healthy\_after_{ik}$ and $Unhealthy\_after_{ik}$ are dummy variables in treatments 2 and 3, indicating whether a food

4 In addition to the variables discussed above, we considered interaction effects between taste and information and between taste, health, and demographics. All interaction terms were statistically insignificant, so dropped them and utilized the more parsimonious model discussed in the main text. Furthermore, note that out experiment relies on a within-subject design, and as such demographics are held constant across treatments for a given individual.

actually fell in quadrants 2 or 4, respectively. The food items in quadrants 1 and 3 are called $Mid - level\ Healthy\_after_{ik}$. The mid-level dummies are dropped such that the effects of the healthy and unhealthy variables are relative to those foods in the intermediate categories.

In this study, the 173 food items were classified into 8 categories: *Cereal, Dairy, Fruit, Meat, Mixed, Snack, Veggie* and *Other*. Cereal products, potatoes, and legumes were included in variable $Cereal_k$ (28 items); dairy products were in $Dairy_k$ (22 items); fruit and fresh processed foods were in $Fruit_k$ (11 items); meat, fish, and eggs were in $Meat_k$ (28 items); mixed dishes like sandwiches and hamburgers were in $Mixed_k$ (14 items); snacks and sweets were in $Snack_k$ (23 items); vegetable and fresh processed foods were in $Veggie_k$ (31 items); and water, coffee, tea, condiments, and oil were in $Other_k$ (16 items). These binary variables take a value of 1 when the associated food item is included in the respective category, and 0 otherwise. For identification, the $Other_k$ variable was dropped so that the effects of other food categories are estimated relative to $Other_k$. The appendix lists all 173 foods, the category in which each was placed, and each food's health classification.

The random utility function consists of a deterministic (V_{ikt}) given in (1) and a stochastic (ε_{ikt}) component. The i^{th} individual's utility of choosing the k^{th} food item in treatment *t* is

(2) $U_{ikt} = V_{ikt} + \varepsilon_{ikt}$,

where V_{ikt} is the systematic utility determined by type of food, perceived taste, healthiness, and price, and ε_{ikt} is a stochastic element which is distributed independently and identically across the *i* individuals, *k* food items, and *t* treatment with a type I extreme value distribution.[5]

5 Following Hausman and McFadden (1984), we tested for violation of the assumption of the independence from irrelevant alternative (IIA). We first estimated the unrestricted model, with all 173 alternative, and then estimated a restricted model, with only 172 alternative (deleting the first option). The Hausman statistic is 2.168, and we fail to reject the null hypothesis, which means IIA assumption holds. Such a test

The probability that the i^{th} individual chooses the k^{th} food item is the conditional logit model

$$(3)\ \ P_{ikt} = \frac{e^{V_{ikt}}}{\sum_{j=1}^{J} e^{V_{ijt}}}.$$

Parameters are estimated by maximizing the log-likelihood function

$$(4)\ \ \log L = \sum_{i=1}^{N} \sum_{k=1}^{J} \sum_{t=1}^{T} q_{ikt} \log(P_{ikt}),$$

where q_{ikt} is the share of total quantity of food purchased by individual i^{th} accounted for by the k^{th} food in treatment t^{th}, and P_{ikt} is defined in (3).[6]

Using the estimated coefficients, we can calculate the willingness-to-pay (WTP) for healthy vs. unhealthy food before and after information. The WTP for healthy vs. mid-level healthy food before information is determined by

$$(5)\ \ WTP_{Healthy\_before} = -\frac{\beta_{Healthy\_before}}{\beta_{price}},$$

where $\beta_{Healthy\_before}$ is the coefficient (marginal utility) associated with the variable $Healthy\_before_{ik}$, and β_{price} is the coefficient associated with the variable $Price_{kt}$. In the same way as (5), we can estimate the WTP for healthy food after receiving information and the WTP

could be re-conducted leaving out any combination of the alternative. For example, when we estimated model with 170 alternative (dropping the first three alterative), the test statistic is actually negative: -5.303, a possibility mentioned by Hausman and McFadden (1984) and discussed by Cheng and Long (2007), but an outcome that would again suggest the IIA assumption is valid.

6 This modeling framework conceptualizes the respondent as making a series of independent choices over each gram of food selected. One could instead conceptualize consumers as maximizing a continuous utility function by choosing quantities of the 173 goods. The appendix shows the results for such an approach where we estimate a series of 173 Tobit models with cross-equation parametric restrictions. The results from this approach are broadly consistent the conditional logit model presented in the main text. The advantage of the conditional logit approach is the ability to calculate the value of information in a theoretic-consistent manner.

for unhealthy food prior to and after information. The WTP for healthy vs. unhealthy food before information is calculated by

(6) $WTP_{Healthy\_before} - WTP_{Unhealthy\_before}$.

Equations (5) and (6) show the tradeoff consumers are willing to make between health and money. Because the taste scale (-5 to +5) is also continuous number, instead of using dollar units, taste units could be used to investigate the relationship between tastiness and healthiness. The willingness-to-give up taste units (WTT) for healthy food relative to the mid-healthy food is

(7) $WTT_{Healthy\_before} = \frac{\beta_{Healthy\_before}}{\beta_{taste}}$,

where β_{taste} is the coefficient (marginal utility) of variable $Taste_{ik}$.

In addition to these calculations, we can also measure the value of information to consumers using the results of the conditional logit model. To determine the value of information (or the cost of imperfect information), Foster and Just (1989) suggest an approach which allows individuals' perception of quality to influence consumption decisions while also allowing true information to influence *ex post* utility. Leggett (2002) applied the Foster and Just (1989) approach to the discrete choice framework used here.

The basic idea behind the approach lies in projecting the welfare loss that would arise if informed consumers were forced to make the same set of choices they did when they were uninformed. We assume the actual nutritional value of each food is constant, but the person's perception of the nutrient content changes after information. As shown by Leggett (2002), the value of the information is

(8) $$\mathrm{CV} = -\frac{1}{\beta_{price}}[\log\left(\sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{t=1}^{T}\exp\left(V_{ikt}^{1*}\right)\right) - \log\left(\sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{t=1}^{T}\exp\left(V_{ikt}^{0*}\right)\right)$$

$$-\sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{t=1}^{T}\pi_{ikt}^{0*}\left(V_{ikt}^{0} - V_{ikt}^{0*}\right)],$$

where $\pi_{ikt}^{0*} = \frac{\exp(V_{ikt}^{0*})}{\sum_i \sum_k \sum_t \exp(V_{ikt}^{0*})}$, *CV* is compensating variation, β_{price} is a coefficient on price, V_{ikt}^{1*} is the *i*th consumer's perception of the *k*th food item's health in treatments 2 and 3 after receiving information, V_{ikt}^{0*} is the *i*th consumer's perception of the *k*th food item's health in treatment 1 before receiving information, V_{ikt}^{0} is the true *k*th food item's health before receiving information in treatment 1, and π_{ikt}^{0*} is the probability of choosing the *k*th food item based on pre-disposed information perception.

Results

Table 1-1 shows how each food type, tastiness, healthiness, and price of food items affects the probability of consumers' food choices. The coefficient for every food type (*Cereal, Dairy, Fruit, Meat, Mixed, Snack,* and *Veggie*) is negative, meaning that the *Other* type of food is preferred to these types. This result might have been obtained because commonly consumed items frequently chosen by a large proportion of consumers, such as water, tea, coffee, and sauce like ketchup and mayonnaise, were classified as *Other*. Aside from *Other*, *Dairy* and *Fruit* were among the most preferred, whereas *Cereal* and *Veggie* were among the least preferred.

Taste has a positive relationship with decision to consume food items. That is, the consumption of tasty foods increases consumers' utility. A one-unit increase in perceived taste of food (on the -5 to +5 scale) increases consumers' utility by 0.534 units. As expected, *Price* has a negative relationship with the probability of consuming food items. Table 1-1-1 indicates that perceived health and health information influence consumers' daily food choices. Prior to receiving information, there is a positive marginal utility for perceived healthy foods ($Healthy\_before_{ik}$) relative to mid-level healthy foods ($Mid - level\ Healthy\_before_{ik}$) from quadrant 1 and 3 in the SAIN/LIM matrix; however, the result is not statistically significant. Conversely, *ceteris paribus,* perceived unhealthy foods ($Unhealthy\_before_{ik}$) yields a negative marginal utility relative to mid-level healthy items ($Mid-level\ Healthy\_before_{ik}$). Upon receiving information pertaining to the healthiness of the 173 food items, the signs of all respective coefficients are the same as the signs of all respective coefficients prior to receiving information, but they are larger in absolute value and statistically significant. Healthy food ($Healthy\_after_{ik}$) has a positive relationship with the decision of purchasing food items and unhealthy food ($Unhealthy\_after_{ik}$) has a negative relationship with the decision to consume food items.

To test if the parameters are statistically different, we calculated each parameter's 95% confidence interval. The respective healthy and unhealthy foods' confidence intervals do not overlap each other. This indicates that although the coefficients have the same sign, they are statistically different, meaning objective information has a certain effect on consumers' food choices. Also, the absolute value of $Healthy\_after_{ik}$ *and* $Unhealthy\_after_{ik}$ are larger than that of $Healthy\_before_{ik}$ *and* $Unhealthy\_before_{ik}$, which means people respond more to objective information than to their beliefs.

Table 1-2 shows the WTP for healthy and unhealthy food. Consumers are willing to pay 0.62€/kg more for healthy food than mid-level healthy food when making decisions based solely on their prior beliefs. When respondents receive objective information regarding the healthiness of food items, their WTP for healthy vs. mid-level healthy food increases to 1.44€/kg. When imperfectly informed, WTP for unhealthy food over mid-level healthy food is -4.99€/kg. This means that consumers are willing to pay an additional 4.99€/kg for mid-level healthy food over unhealthy food. Additionally, the results indicate that consumers are willing to pay 14.24€/kg for mid-level healthy food as opposed to unhealthy food when perfect information is received. The results suggest a type of loss aversion in that losses (unhealthy food) have a larger impact than gains (healthy food). Table 1-2 also indicates how much consumers are willing to pay for healthy food rather than unhealthy food. Prior to information, they are willing to pay 5.62€/kg more for healthy food than unhealthy food. After the nutrient information, the WTP for healthy food rather than unhealthy food is almost three times larger at 15.68€/kg. This result suggests if people could access precise healthiness information about foods, they are willing to pay more for healthy foods.

When it comes to perceived taste of food, people are willing to pay 4.33€/kg more for a one-unit increase on the -5 to +5 taste scale.[7] To put this number in perspective, the appendix lists the average taste rating given to all 173 food items. Most items had a mean rating above zero. The highest rated items on average were items like tomatoes (+4.1), green salad (+4), and zucchini (+3.9). The lowest rated items on average included cheese spread (-0.2) and Orangina light (-1.9). Moving from one of the lower to higher rated items would induce a four-point change in the taste scale associated with a change in economic value of 4.33*4 = 17.32€/kg (see Table A1-3).

It is also possible to calculate how much taste unit people are willing to give up to get healthy food rather than unhealthy food in both informed and uninformed situation. Before consumers receive the nutrient information, they are willing to give up 1.29 taste units to have a healthy food rather than an unhealthy food on the -5 to +5 taste scale. After provided information, the taste tradeoff is 3.61 units to have a healthy food rather than unhealthy food. That is, when

7 $WTP_{taste} = -\frac{\beta_{taste}}{\beta_{price}}$, where β_{taste} is the coefficient of taste variable.

consumers receive the nutrient information, they are more willing to sacrifice taste units for healthiness.

Plugging the estimates in Table 1-1 into equation (8), we can estimate the value of information. Results indicate that given the average quantity of food chosen per day in the experiment, the value of LIM/SAIN quadrant nutrient information to consumers is €0.98/family/day. The 95% confidence lower limit and upper limit are 0.872 and 1.324, respectively. When we consider other value of information estimates that have used the Leggett (2002) approach, €0.98/day is a sensible value. Ellison et al. (2014) measured the value of the numeric calorie labels and the value of the symbolic calorie label, which were estimated at \$0.03/dinner/meal and \$0.13/dinner/meal, respectively. Brooks and Lusk (2010) estimated a value of mandatory labeling for milk from cloned cattle at \$0.19 per time the consumer chooses to buy milk. Hu, Veeman, and Adamowicz (2005) estimated the value of genetically modified food labeling policy. Their estimates ranged from \$0 to \$0.15 per time the consumer chooses bread. Klain et al. (2014) used two different approaches to measure the value of country of origin information for beef and pork, and found values that ranged from \$1.36 to \$2.15 per choice occasion. Lastly, Tiesl et al. (2001) estimated the value of nutritional information of 6 food items, and found that the milk's value of information is the highest—\$0.434/month—the peanut butter's value of information is the second highest—\$0.336/month—and the lowest value of information is cream cheese—\$0.002/month. Because these studies utilize different units, different information, and food items, it is difficult to compare their values with our values directly. However, our estimate of €0.98/day does not seem out of line with these previous estimates, particularly because our estimate is a value of information over all food products eaten during a day.

Summary and Conclusion

In this study, we found that nutrient information conveyed through simple indices influences consumers' grocery choices. Nutrient information increases willingness-to-pay (WTP) for healthy food and decreases WTP for unhealthy food. The added certainty provided by objective nutrient information increased the marginal WTP for healthy food. Moreover, there is a sort of loss aversion at play in that WTP for healthy vs. neutral food is lower than WTP for neutral vs. unhealthy food, and this loss aversion increases with information. The result suggests that a label design with emphasis on negative nutrient information could be more influential in improving the healthfulness of consumers' food choice than one that focuses on positive nutrient information. In fact, the U.S. FDA has changed the nutrient facts label in 2016, and they seem to focus on highlighting negative information by making caloric and added-sugar content more prominent.

This study estimated the value of the nutrient index information at €0.98/family/day. The advantage of our approach is that the value of information reflects choices over a larger number of possible foods and represents an aggregate value over the whole day. Previous attempts to provide a monetary estimate of the value of nutritional information have tended to focus on a single product or product category. One downside of our approach is that it likely represents an upper-bound to the value of information. The value of information is directly tied to the change in choices that occur as a result of information provision, and our experiment focused people's attention on this particular issue. In a real life grocery setting, it would be difficult to get consumers to invest the same level of cognitive resources in investigating the healthiness of each and every food item they might consider. Nonetheless, it is useful when considering the costs and benefits of policies related to nutrient labeling to have bounds on possible benefits.

References

Balcombe, K., I. Fraser, and S. Di Falco. 2010. "Traffic Lights and Food Choice: A Choice Experiment Examining the Relationship between Nutritional Food Labels and Price." *Food Policy* 35(3):211-220.

Brooks, K., and J.L. Lusk. 2010. "Stated and Revealed Preferences for Organic and Cloned Milk: Combining Choice Experiment and Scanner Data" *American Journal of Agricultural Economics* 92(4):1229-1241.

Chang, J.B., J.L. Lusk, and F.B. Norwood. 2009. "How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behavior?" *American Journal of Agricultural Economics* 19(2):518-534.

Christopher, G.L. 2002. "Environmental Valuation with Imperfect Information: The Case of the Random Utility Model." *Environmental and Resource Economics* 23(3):345-355.

Darmon, N., F. Vieux, M. Maillot, J.L. Volatier, and A. Martin. 2009. "Nutrient Profiles Discriminate between Foods According to Their Contribution to Nutritionally Adequate Diets: A Validation Study Using Linear Programming and the SAIN, LIM System." *American Journal of Clinical Nutrition* 89(4):1227-1236.

Derby, B.M., and A.S. Levy. 2001. "Do Food Labels Work? Gauging the Effectiveness of Food Labels Pre- and Post-NLEA." In P.N. Bloom and G.T. Gundlach, eds. *Handbook of Marketing and Society*. Thousand Oaks CA: Sage, pp.372-398.

Drichoutis, A.C., Jr.R.M. Nayga, P. Lazaridis, J.L. Lusk, J. Roosen, and J.F. Shogren. 2011. "Nutritional Labeling." *The Oxford Handbook of the Economics of Food Consumption and Policy*:520-545.

Drichoutis, A.C., P. Lazaridis, and Jr.R.M. Nayga. 2006. "Consumers' Use of Nutritional Labels: A Review of Research Studies and Issues." *Academy of Marketing Science* 2006(9):1-22.

Drichoutis, A.C., P. Lazaridis, and Jr.R.M. Nayga. 2005. “Nutrition Knowledge and Consumer Use of Nutritional Food Labels” *European Review of Agricultural Economics* 32(1):93-118.

Ellison, B., J.L. Lusk, and D. Davis, 2014. “The Impact of Restaurant Calorie Labels on Food Choice: Results from a Field Experiment.” *Economic Inquiry* 52(2):666-681.

Feunekes, G.I.J., I.A. Gortemaker, A.A. Willems, R. Lion, M. Kommer. 2008. “Front-of-Pack Nutrition Labeling: Testing Effectiveness of Different Nutrient Labeling Formats Front-of-Pack in Four European Countries.” *Appetite* 50(1):57-70.

Food and Drug Administration. 2014. “Food Labeling: Revision of the Nutrition and Supplement Facts Labels.” *The Daily Journal of the United States Government* 79(41):11880-11987. Silver Spring MD.

Foster, W., and R.E. Just. 1989. “Measuring Welfare Effects of Product Contamination with Consumer Uncertainty.” *Journal of Environmental Economics and Management* 17(3):266-283.

Garcia, A., M. Loureiro, and Jr.R. M. Nayga. 2007. “Do Consumers Perceive Benefits from the Implementation of a EU Mandatory Nutritional Labeling Program?” *Food Policy* 32(2):160-174.

Garretson, J.A., and S. Burton. 2000. “Effects of Nutrition Facts Panel Values, Nutrition Claims, and Health Claims on Consumer Attitudes, Perceptions of Disease-Related Risks, and Trust.” *Journal of Public Policy & Marketing* 19(2):213-227.

Greene, W. H. 2002. *Econometric Analysis*. 5th. ed. Englewood Cliffs NJ: Prentice-Hall.

Grunert, K.G., and J.M. Wills. 2007. “A Review of European Research on Consumer Response to Nutrition Information on Food Labels.” *Journal of Public Health* 15(5):385-399.

Hausman, J., and D. McFadden. 1984. “Specification Tests for the Multinomial Logit Model.” *Econometrica: Journal of the Econometric Society*:1219-1240.

Hu, W., M.M. Veeman, and W.L. Adamowicz. 2005. “Labeling Genetically Modified Food: Heterogeneous Consumer Preferences and the Value of Information” *Canadian Journal of Agricultural Economics/Revenue Canadienne d’Agtoeconomie* 53(1):83-102.

Ippolito, P.M., and A.D. Mathios. 1991. “Health Claims in Food Marketing: Evidence on Knowledge and Behavior in the Cereal Market.” *Journal of Public Policy & Marketing* 10(1):15-32.

Kahneman, D., and A. Tversky. 2000. “Choices, Values, and Frames.” *Cambridge University Press.*

Kiesel, K., and S.B.V. Boas. 2013. “Can Information Costs Affect Consumer Choice? Nutritional Labels in a Supermarket Experiment.” *International Journal of Industrial Organization* 31(2):153-163.

Klain, T.J., J.L. Lusk, G.T. Tonsor, and T.C. Schroeder. 2014. “An Experimental Approach to Valuing Information” *Agricultural Economics* 45(5):635-648.

Kreuter, M.W., L.K. Brennan, D.P. Shriff, and S.N. Lukwago. 1997. “Do Nutrition Label Readers Eat Healthier Diets? Behavioral Correlates of Adults’ Use of Food Label.” *American Journal of Preventive Medicine* 13:277-283

Kristin, K., and S.B.V. Boas. 2013. “Can Information Costs Affect Consumer Choice? Nutritional Labels in a Supermarket Experiment.” *International Journal of Industrial Organization* 31(2):153-163.

Leggett, C.G. 2002. “Environmental Valuation with Imperfect Information the Case of the Random Utility Model.” *Environmental and Resource Economics* 23(3):343-355.

Lin, C.J., J. Lee, and S.T. Yen. 2004. “Do Dietary Intakes Affect Search for Nutrient Information on Food Labels?” *Social Science & Medicine* 59(9):1955-1967.

Mai, R., and S. Hoffmann. 2014. “How to Combat the Unhealthy=Tasty Intuition: The Influencing Role of Health Consciousness.” *Journal of Public Policy & Marketing*:in press.

Mario, F.T., N.E. Bockstael, and A. Levy. 2001. “Measuring the Welfare Effects of Nutrition Information.” *American Journal of Agricultural Economics* 83(1):133-149.

McFadden, D. 1973. “Conditional Logit Analysis of Qualitative Choice Behavior” *Frontiers in Econometrics*:105-142.

Muller, L., A. Lacroix, J.L. Lusk, and B. Ruffieux. 2016. “Distributional Impacts of Fat Taxes and Thin Subsidies.” *The Economic Journal*: in press.

Rabin, M., and J.L. Schrag. 1999. “First Impressions Matter: A Model of Confirmatory Bias.” *Quarterly Journal of Economics* 114(1): 37-82.

Raghunathan, R., R.W. Naylor, and W.D. Hoyer. 2006. “The Unhealthy=Tasty Intuition and Its Effects on Taste Inferences, Enjoyment, and Choice of Food Products.” *Journal of Marketing* 70(4):170-184.

Roberto, C.A., M.A. Bragg, M.B. Schwartz, M.J. Seamans, A. Musicus, N. Novak, and K.D. Brownell. 2012. “Facts up Front Versus Traffic Light Food Labels: A Randomized Controlled Trial.” *American Journal of Preventive Medicine* 43(2):134-141.

Sacks, G., M. Rayner, and B. Swinburn. 2009. “Impact of Front-of-Pack ‘Traffic-Light’ Nutrition Labeling on Consumer Food Purchases in the UK.” *Health Promotion International* 24(4):344-352.

Smith, J.P. 2004. “Unraveling the SES: Health Connection.” *Population and Development Review* 30:108-132.

Smith, T.G. 2004. “The McDonald’s Equilibrium: Advertising, Empty calories, and the Endogenous Determination of Dietary Preferences.” *Social Choice and Welfare* 23(3):383-413.

Teisl, M.F., N.E. Bockstael, and A.S. Levy. 2001. “Measuring the Welfare Effects of Nutrition Information.” *American Journal of Agricultural Economics* 83(1):133-149.

Tepper, B.J., and A.C. Trail. 1998. “Taste or Health: A Study on Consumer Acceptance of Corn Chips.” *Food Quality and Preference* 9(4):267-272

Timothy, R., K. Yonezawa, and S. Winter. 2013. “Cross-Category Effects and Private

Labels." *European Review of Agricultural Economics* June:1-30.

Wansink, B. 2004. "Environmental Factors that Increase the Food Intake and Consumption Volume of Unknowing Consumers." *Annu. Rev. Nutr*. 24:455-479.

Weaver, D., and M. Finke. 2003. "The Relationship between the Use of Sugar Content Information on Nutrition Labels and the Consumption of Added Sugars" *Food Policy* 28(3):213-219.

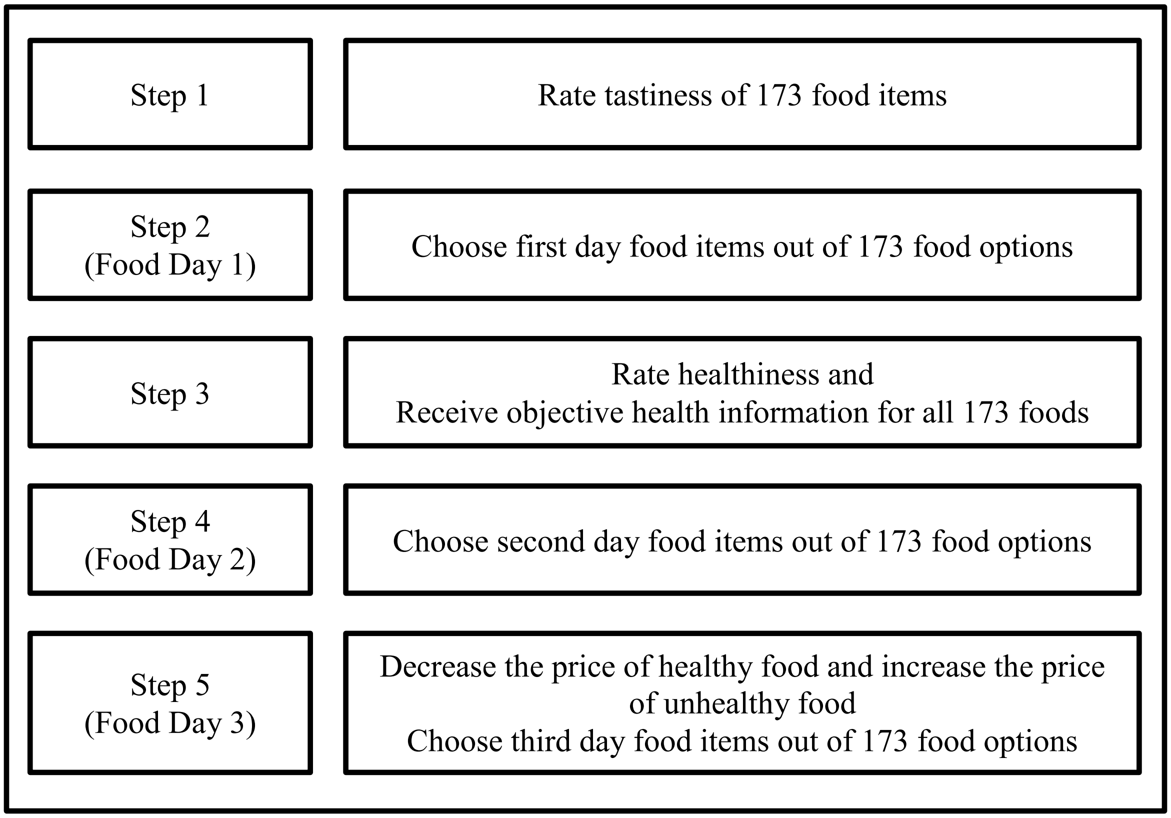


Figure 1-1. Steps in the experiment

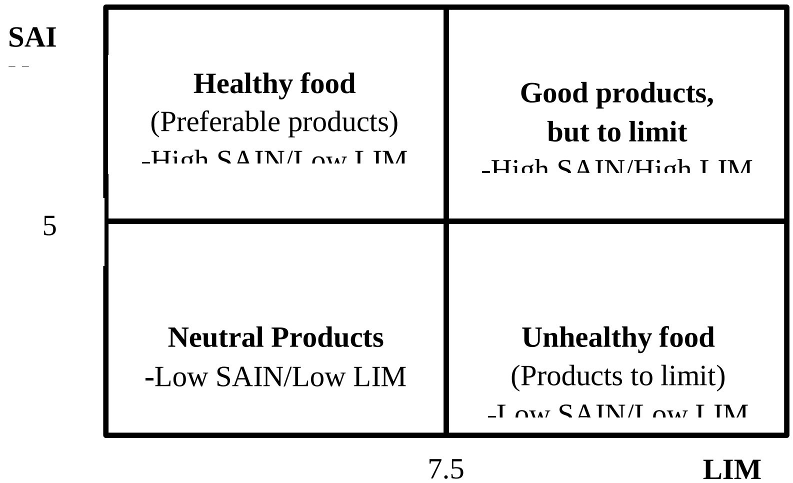


Figure 1-2. Four categories of SAIN and LIM score

Table 1-1. Conditional logit estimates

Variable	Estimate
Cereal	-1.421**(0.187)
Dairy	-1.080**(0.168)
Fruit	-1.112**(0.205)
Meat	-1.411**(0.225)
Mixed	-1.294**(0.332)
Snack	-1.136**(0.278)
Veggie	-1.673**(0.167)
Taste	0.534**(0.043)
Healthy_before	0.077(0.050)
Unhealthy_before	-0.615*(0.298)
Healthy_after	0.178**(0.038)
Unhealthy_after	-1.753**(0.316)
Price	-0.123**(0.024)

Notes: N=387. Standard errors in parentheses. An * denotes significance at the 5% level, and ** denotes significance at the 1% level.

Table 1-2. Willingness-to-pay for healthy and unhealthy food (€/kg consumed) and Willingness-to-give up taste for healthy and unhealthy food (taste units)

Willingness-to-Pay	Before information	After information
Healthy vs. neutral	0.625€ (0.433)	1.442€ (0.444)
Unhealthy vs. neutral	-5.000€ (2.642)	-14.243€ (3.881)
Healthy vs. unhealthy	5.624€ (2.618)	15.685€ (4.084)
Taste tradeoff	Before information	After information
Healthy vs. neutral	0.144 taste units (0.095)	0.332 taste units (0.077)
Unhealthy vs. neutral	-1.152 taste units (0.568)	-3.282 taste units (0.651)
Healthy vs. Unhealthy	1.296 taste units (0.550)	3.615 taste units (0.651)

Notes: Standard errors in parentheses.

Appendix

One downside of the CL model above is that it does not take into consideration the fact that respondents could choose multiple items. Our implementation of the model analyzes the share of purchases allocated to different items, and as such it imagines a consumer making a series of many (independent) choices about whether or not to buy a gram of each product. Because this may not match the approach consumers actually utilized to make their food purchase, we consider another econometric approach that is more flexible, but admittedly ad hoc in the sense that the estimated demands may not integrate back to a well-defined utility function.

To investigate the robustness of our results, we estimate a series of 173 Tobit models with cross-equation parametric restrictions, where the dependent variables are the quantities of each good purchased. The Tobit model is used because the dependent variable is censored at zero. The likelihood function of a general censored regression model is

(9) $L = \prod_{i=1}^{N} \prod_{t=1}^{T} \{\frac{1}{\sigma} \emptyset(\frac{y_{ikt}-X_{ikt}\beta}{\sigma})\}^{d_{ikt}} \{\Phi(\frac{-X_{ikt}\beta}{\sigma})^{(1-d_{ikt})}\}$,

where $X_{ikt}\beta = \beta_1 cereal_k + \beta_2 dairy_k + \beta_3 fruit_k + \beta_4 meat_k + \beta_5 mixed_k + \beta_6 snack_k + \beta_7 veggie_k + \beta_8 other_k + \beta_9 t_{ik} + \beta_{10} Healthy\_before_{ikt} + \beta_{11} Unhealthy\_before_{ikt} + \beta_{12} Healthy\_after_{ikt} + \beta_{13} Unhealthy\_after_{ikt} + \beta_{14} price_{kt}$,

y_{ikt} is the dependent variable consisting of the quantity of the k^{th} food purchased by individual i^{th} in treatment t, $\emptyset$ is the standard normal density function, Φ is the standard normal cumulative density function, and d_{ikt} is the dummy variable which takes 1 for $y_{ikt} > 0$ and 0 for $y_{ikt} = 0$.

Table A1-1 reports the estimated coefficients. There constants associated with each food type are negative, indicating the fact that there are many observations with zero purchases. However, like the conditional logit results presented in the main text, the constant on *Other* is higher than on the other food categories. As in the conditional logit, the price effect is negative (the demand curves are downward sloping) and the taste effect is positive (tastier foods are in higher demand). Table A.1-1 also shows that in both cases, before receiving the information and after receiving the information, there is a positive relationship between healthy food and the probability of purchasing quantity and a negative relationship between unhealthy food and food consuming decision.

We can also report a measure somewhat similar to WTP. In particular, we ask what price difference between two items (with different health scores) would generate the same quantity purchased. Quantity-equivalent prices for healthy and unhealthy food from the Tobit model are reported in Table A1-2. Consumers are willing to pay 1.33€/kg more for healthy food than mid-level healthy food and keep the same purchasing quantity when they do not have perfect information. After consumers receive perfect nutrient information, crossed quantity-equivalent prices for healthy food is increased by 2.46€/kg when they keep the same consuming quantity. If people receive perfect information, they are willing to pay more for healthy food than mid-level healthy food.

When individuals do not have perfect information, they are willing to pay an additional 3.99€/kg for mid-level healthy food as opposed to unhealthy food to keep their food purchasing quantity decision. Also, in perfectly informed situations, crossed quantity-equivalent prices for unhealthy food is -7.36€/kg, which is almost twice as large as crossed quantity-equivalent prices of imperfectly informed situations. Thus, when people receive perfect nutrient information, they are willing to pay more to avoid unhealthy food.

Lastly, Table A1-2 also describes how much more people are willing to pay for healthy food rather than unhealthy food in both imperfectly informed situations and perfectly informed situations. When consumers do not have perfect nutrient information, they are willing to pay 5.33€/kg more for healthy food than unhealthy food. After they receive the nutrient information, crossed quantity-equivalent prices for healthy food rather than unhealthy food is 9.83€/kg. Therefore, we can say that if the nutrient information is provided to people, they prefer healthy food to unhealthy food.

Table A1-1 Tobit model parameter estimate of each attributes

Variable	Estimate
Cereal	-504.380**(8.647)
Dairy	-437.130**(8.201)
Fruit	-454.960**(9.747)
Meat	-556.990**(10.334)
Mixed	-612.140**(13.502)
Snack	-539.660**(10.963)
Veggie	-529.610**(9.240)
Other	-288.160**(7.109)
Taste	47.590**(1.123)
Healthy_before	9.630**(1.640)
Unhealthy_before	-28.755**(8.458)
Healthy_after	17.764**(1.352)
Unhealthy_after	-53.030**(6.094)
Price	-7.199**(0.505)
Sigma	265.830**(2.818)

Notes: N=387. Standard errors in parentheses. An * denotes significance at the 5% level, and ** denotes significance at the 1% level.

Table A1-2 Crossed quantity-equivalent prices for healthy and unhealthy food from Tobit model

Crossed quantity-equivalent prices	Before information	After information
Healthy vs. neutral	1.338€/kg (0.250)	2.468€/kg (0.263)
Unhealthy vs. neutral	-3.994€/kg (1.208)	-7.366€/kg (1.001)
Healthy vs. unhealthy	5.332€/kg (1.215)	9.834€/kg (1.079)

Table A1-3 Tastiness rating of 173 food items

Rank	Food item	Category	Healthiness	Mean Taste	Std Dev
1	Stuffed tomatoes	Vegetables, Fresh & Processed	Healthy	4.152	1.477
2	Tap water	Others	Neutral	4.000	1.532
3	Green salad	Vegetables, Fresh & Processed	Healthy	3.904	1.646
4	Zucchini	Vegetables, Fresh & Processed	Healthy	3.674	1.850
5	Baguette	Cereals, potatoes, Legumes	Neutral	3.669	2.051
6	Clementine	Fruits, Fresh & Processed	Healthy	3.643	1.770
7	Fresh fruit salad	Fruits, Fresh & Processed	Healthy	3.610	2.014
8	Pasta	Cereals, Potatoes, Legumes	Neutral	3.607	1.657
9	French bean	Vegetables, Fresh & Processed	Healthy	3.491	1.667
10	Carrot	Vegetables, Fresh & Processed	Healthy	3.457	1.940
11	Smoked salmon	Meat, Fish & Eggs	Good but limited	3.455	2.492
12	Farmhouse bread	Cereals, Potatoes, Legumes	Neutral	3.434	1.709
13	Shrimp	Meat, Fish & Eggs	Healthy	3.421	2.353
14	White rice	Cereals, Potatoes, Legumes	Neutral	3.339	1.669
15	Grated carrot	Vegetables, Fresh & Processed	Healthy	3.318	2.026
16	Ratatouille	Vegetables, Fresh & Processed	Healthy	3.214	2.272
17	Roasted chicken legs	Meat, Fish & Eggs	Healthy	3.214	1.897
18	Orange	Fruits, Fresh & Processed	Healthy	3.160	2.220
19	Whole bread	Cereals, Potatoes, Legumes	Healthy	3.119	2.045
20	Spinach	Vegetables, Fresh & Processed	Healthy	3.103	2.400
21	Grilled beef steak	Meat, Fish & Eggs	Healthy	3.057	2.281
22	Mashed potatoes	Cereals, Potatoes, Legumes	Healthy	3.054	2.024
23	Dark chocolate	Snack & Sweets	Unhealthy	3.039	2.501
24	Cheese pizza	Mixed Dishes	Unhealthy	3.028	2.313
25	Poivron	Vegetables, Fresh & Processed	Healthy	2.990	2.656
26	Squeezed orange juice	Fruits, Fresh & Processed	Healthy	2.961	2.214
27	Unsalted chips	Cereals, Potatoes, Legumes	Unhealthy	2.961	2.307
28	Flan	Snack & Sweets	Unhealthy	2.935	2.251

29	Eggplant	Vegetables, Fresh & Processed	Healthy	2.917	2.384
30	Ice cream	Snack & Sweets	Unhealthy	2.915	2.220
31	Apple	Vegetables, Fresh & Processed	Healthy	2.884	2.345
32	Crepe	Snack & Sweets	Unhealthy	2.879	2.580
33	Tabbouleh	Cereals, Potatoes, Legumes	Healthy	2.876	2.277
34	Banana	Fruits, Fresh & Processed	Healthy	2.860	2.783
35	Cucumber	Vegetables, Fresh & Processed	Healthy	2.853	2.561
36	Lasagna	Mixed Dishes	Good but limited	2.848	2.679
37	Jam	Fruits, Fresh & Processed	Unhealthy	2.832	1.946
38	Kiwi	Vegetables, Fresh & Processed	Healthy	2.796	2.767
39	Sherbet	Snack & Sweets	Unhealthy	2.755	2.416
40	Lens	Cereals, Potatoes, Legumes	Healthy	2.747	2.344
41	Croissant	Cereals, Potatoes, Legumes	Unhealthy	2.747	2.403
42	Boiled potatoes	Cereals, Potatoes, Legumes	Healthy	2.731	2.124
43	Sweet apple sauce	Dairies	Neutral	2.726	2.536
44	Chocolate croissant	Cereals, Potatoes, Legumes	Unhealthy	2.700	2.610
45	Avocado	Vegetables, Fresh & Processed	Healthy	2.698	2.931
46	Grated Swiss cheese	Dairies	Good but limited	2.674	2.306
47	Fresh vegetable soup	Vegetables, Fresh & Processed	Healthy	2.669	2.706
48	Lemon yellow	Fruits, Fresh & Processed	Healthy	2.664	2.329
49	Pear	Dairies	Healthy	2.664	2.610
50	Beefsteak	Meat, Fish & Eggs	Good but limited	2.633	2.300
51	Chocolate mousse	Snack & Sweets	Unhealthy	2.633	2.621
52	Canned tuna in brine	Meat, Fish & Eggs	Healthy	2.584	2.258
53	Coffee	Others	Healthy	2.568	3.020
54	Plain omelet	Meat, Fish & Eggs	Healthy	2.566	2.288
55	Salami	Meat, Fish & Eggs	Unhealthy	2.543	3.113
56	Cured ham	Mixed Dishes	Good but limited	2.509	3.064
57	Emmental cheese	Dairies	Good but limited	2.506	2.314
58	Mineralized water	Others	Neutral	2.494	2.532
59	Tea	Others	Healthy	2.494	2.667

60	Carbonated mineral water	Others	Neutral	2.494	2.532
61	Tomato salad	Vegetables, Fresh & Processed	Healthy	2.483	2.860
62	Hard boiled egg	Meat, Fish & Eggs	Good but limited	2.439	2.355
63	Crème fraiche	Dairies	Unhealthy	2.439	2.087
64	Milk chocolate	Snack & Sweets	Unhealthy	2.429	2.750
65	Cooked ham	Mixed Dishes	Good but limited	2.419	2.977
66	Plain yogurt	Dairies	Healthy	2.382	2.829
67	Hake	Meat, Fish & Eggs	Healthy	2.382	2.498
68	Tin	Meat, Fish & Eggs	Healthy	2.377	3.194
69	Potato salad	Cereals, Potatoes, Legumes	Neutral	2.377	2.642
70	Brioche	Cereals, Potatoes, Legumes	Unhealthy	2.354	2.415
71	Cod	Meat, Fish & Eggs	Healthy	2.336	2.490
72	Leek tart	Mixed Dishes	Unhealthy	2.320	2.725
73	Goat soft cheese	Dairies	Unhealthy	2.310	3.316
74	Roast breast of duck	Meat, Fish & Eggs	Good but limited	2.289	3.007
75	Peanut oil	Others	Unhealthy	2.284	1.908
76	Oil	Others	Good but limited	2.284	1.908
77	Apple juice	Fruits, Fresh & Processed	Neutral	2.274	2.492
78	Tiramisu	Snack & Sweets	Unhealthy	2.271	3.015
79	Couscous	Mixed Dishes	Unhealthy	2.266	3.101
80	Hazelnut soft margarine	Snack & Sweets	Unhealthy	2.209	3.386
81	Unsalted butter	Dairies	Unhealthy	2.181	2.093
82	Margarine	Dairies	Unhealthy	2.181	2.093
83	Vinaigrette	Others	Unhealthy	2.176	2.526
84	Soft corn	Vegetables, Fresh & Processed	Neutral	2.173	2.625
85	Trout	Meat, Fish & Eggs	Healthy	2.160	2.878
86	Éclair	Snack & Sweets	Unhealthy	2.134	2.873
87	Beef bourguignon	Meat, Fish & Eggs	Unhealthy	2.134	2.674
88	Mustard	Others	Good but limited	2.119	2.429
89	Lamb chop	Meat, Fish & Eggs	Good but limited	2.111	2.873
90	Quiche lorraine	Mixed Dishes	Unhealthy	2.103	2.994
91	Frozen apple hazelnut	Cereals, Potatoes, Legumes	Neutral	2.090	2.902

92	Fruit yogurt	Dairies	Good but limited	2.088	2.701
93	Fish stick	Meat, Fish & Eggs	Unhealthy	2.068	2.170
94	Salt	Others	Unhealthy	2.049	2.176
95	Sugar	Others	Unhealthy	2.044	2.588
96	Salted potato chips	Cereals, Potatoes, Legumes	Unhealthy	2.034	2.711
97	Whiting	Meat, Fish & Eggs	Healthy	1.982	2.616
98	Reblochon	Dairies	Good but limited	1.956	3.248
99	Hazelnut	Cereals, Potatoes, Legumes	Neutral	1.933	2.612
100	Camembert	Dairies	Good but limited	1.928	3.097
101	Croque-monsieur	Mixed Dishes	Unhealthy	1.928	3.048
102	Chocolate bar	Vegetables, Fresh & Processed	Unhealthy	1.897	2.837
103	Mixed vegetables	Vegetables, Fresh & Processed	Healthy	1.868	2.506
104	Drinking chocolate	Snack & Sweets	Unhealthy	1.858	3.330
105	UHT skimmed milk	Dairies	Good but limited	1.858	3.330
106	UHT semi-skimmed milk	Dairies	Healthy	1.858	3.330
107	UHT whole milk	Dairies	Healthy	1.858	3.330
108	Fresh garlic	Vegetables, Fresh & Processed	Healthy	1.837	2.895
109	Swiss cheese & ham sandwich	Mixed Dishes	Good but limited	1.778	3.123
110	Rabbit	Meat, Fish & Eggs	Healthy	1.755	3.229
111	Madeleine	Snack & Sweets	Unhealthy	1.747	2.603
112	Almond	Cereals, Potatoes, Legumes	Neutral	1.744	2.698
113	Herb tea	Others	Healthy	1.744	2.980
114	Peanut	Snack & Sweets	Unhealthy	1.711	2.612
115	Coalfish	Meat, Fish & Eggs	Healthy	1.693	2.683
116	Pepper	Others	Healthy	1.669	2.607
117	Caramel tart	Snack & Sweets	Unhealthy	1.638	2.918
118	Rusk	Cereals, Potatoes, Legumes	Neutral	1.581	2.551
119	Diluted fruit syrup	Snack & Sweets	Unhealthy	1.545	3.145
120	Cottage pie	Mixed Dishes	Unhealthy	1.506	3.020
121	Cheese biscuit	Cereals, Potatoes, Legumes	Unhealthy	1.481	2.394
122	Chewing gum	Snack & Sweets	Unhealthy	1.481	2.826
123	Soft white cheese	Dairies	Healthy	1.452	3.025
124	Roast pork	Meat, Fish & Eggs	Healthy	1.452	2.917

125	Candy	Snack & Sweets	Unhealthy	1.395	2.954
126	Hamburger	Mixed Dishes	Unhealthy	1.388	3.390
127	Onion	Vegetables, Fresh & Processed	Healthy	1.370	2.941
128	Grape juice	Fruits, Fresh & Processed	Neutral	1.341	2.996
129	Frozen french bean	Vegetables, Fresh & Processed	Healthy	1.331	2.834
130	Roquefort	Dairies	Good but limited	1.331	3.645
131	Apricot nectar	Fruits, Fresh & Processed	Unhealthy	1.326	3.067
132	Bun	Cereals, Potatoes, Legumes	Unhealthy	1.320	2.856
133	Pamplemousse	Dairies	Healthy	1.313	3.521
134	Petits pois	Vegetables, Fresh & Processed	Healthy	1.261	2.942
135	Bifidus plain yogurt	Dairies	Healthy	1.183	2.912
136	Sandwich bread	Cereals, Potatoes, Legumes	Unhealthy	1.145	2.865
137	Tomato sauce	Others	Good but limited	1.065	2.728
138	Butter cookies	Cereals, Potatoes, Legumes	Unhealthy	1.005	3.007
139	Chocolate biscuit	Cereals, Potatoes, Legumes	Unhealthy	0.990	2.998
140	Chocolate cream dessert	Snack & Sweets	Good but limited	0.941	3.233
141	Dried dates	Vegetables, Fresh & Processed	Neutral	0.928	3.338
142	Diced mixed vegetables	Vegetables, Fresh & Processed	Healthy	0.910	2.961
143	Mayonnaise	Others	Unhealthy	0.899	2.920
144	Cauliflower	Vegetables, Fresh & Processed	Healthy	0.853	2.897
145	Cheeseburger	Mixed Dishes	Unhealthy	0.848	3.602
146	Vegetable soup	Vegetables, Fresh & Processed	Healthy	0.827	2.976
147	Sandwich kebab	Mixed Dishes	Unhealthy	0.739	3.499
148	Fish soup	Meat, Fish & Eggs	Healthy	0.682	3.499
149	Tomato meat sauce ravioli	Meat, Fish & Eggs	Unhealthy	0.664	3.251
150	Farmhouse pate	Meat, Fish & Eggs	Good but limited	0.661	3.322
151	Dried figs	Vegetables, Fresh & Processed	Healthy	0.659	3.664
152	Coca cola	Snack & Sweets	Unhealthy	0.630	3.665
153	Cooked white cabbage	Vegetables, Fresh & Processed	Unhealthy	0.571	3.459
154	Slightly salted butter	Dairies	Unhealthy	0.568	3.057

155	Quenelle	Meat, Fish & Eggs	Unhealthy	0.558	3.375
156	Chipolata	Meat, Fish & Eggs	Unhealthy	0.491	3.340
157	Sardine in oil	Meat, Fish & Eggs	Good but limited	0.475	3.301
158	Pineapple in syrup	Fruits, Fresh & Processed	Good but limited	0.429	3.036
159	Potato gratin	Mixed Dishes	Unhealthy	0.385	3.325
160	Plain corn flakes	Cereals, Potatoes, Legumes	Good but limited	0.377	3.004
161	Chocolate cereal	Cereals, Potatoes, Legumes	Good but limited	0.354	3.145
162	Raisins	Cereals, Potatoes, Legumes	Neutral	0.331	3.298
163	Flageolet bean	Vegetables, Fresh & Processed	Healthy	0.313	3.229
164	Sugar cereal	Cereals, Potatoes, Legumes	Good but limited	0.302	3.108
165	Ketchup	Others	Good but limited	0.199	3.366
166	Exotic fruits with dried seed	Fruits, Fresh & Processed	Unhealthy	0.194	3.335
167	Frankfurter	Meat, Fish & Eggs	Unhealthy	-0.023	3.133
168	Cheese spread	Dairies	Unhealthy	-0.202	3.488
169	Orangina	Snack & Sweets	Unhealthy	-0.243	3.571
170	Coca cola light	Snack & Sweets	Neutral	-0.969	3.703
171	Orangina light	Snack & Sweets	Healthy	-1.907	3.252
172	Barre minceur bar	Vegetables, Fresh & Processed	Healthy	-2.008	2.869
173	Sweetener	Snack & Sweets	Unhealthy	-2.323	2.822

CHAPTER II

CONSUMER-ORIENTED FOODS CLUSTER USING CROSS-NATIONL DATA

Introduction

Understanding consumers' purchasing motivations drives much of the research in modern retailing. As such, widely used category management (CM) standards have evolved to center on shopping behavior (Dudlicek, 2016). Karolefski (2016) summarized the trend in retailing by stating, "Supermarkets are facing tidal pressures from shoppers who want their stores to evolve with their tastes and habits, so businesses need to resist the urge to remain complacent" (p.2). Despite 20 year old arguments that CM should focus on delivering consumer value (e.g., Joint Industry Project on Efficient Consumer Response, 1995), Holweg, Schnedlitz and Teller (2009) argue that the CM process does not sufficiently consider empirical evidence based on consumer-oriented data.

To address this problem, the paper analyzes consumers' perceptions of the taste, health, and affordability of a wide variety of food products to determine how different foods are categorized from the consumers' perspective. Perceived taste, health, and expense concerning foods are chosen as key factors driving potential categorizations as previous literature has identified these factors to be key drivers of consumers' purchasing behavior (Glanz et al., 1998; Lusk and Briggeman, 2009; Zakowska-Biemans, 2011).

Desrochers and Nelson (2006) suggest improvements in management and marketing strategies by using two consumer behavior concepts—category-dependence effects and carryover effects—as a supplement to point-of-purchase scanner information. Category-dependence effects indicate that consumers' preference for a product's attribute could be affected by where the product is categorized. Carryover effects represent the importance of sequential exposures to a product class. For example, an Asian brand name of tofu might have a strong advantage if it is seen first in the Oriental food category, but this advantage would be absent if the Oriental category is seen after the Dairy category. Though Desrochers and Nelson (2006) position consumer behavior concepts for the first step of the effective CM, the authors do not provide the specific assortment examples. Moreover, their empirical work involved an experiment for only two products, Nachos and Tofu. However, our research provides specific examples of classification for a wide variety of food items rather than focusing on only a few. Identifying an efficient assortment not only has the potential to increase sales, margins, and market shares, but also reduces costs for the retailer by implementing the appropriate strategy, promotion, and marketing.

Globalization and the increasing number of multinational companies motivate the necessity of cross-national research. What "works" in one part of the world might not be applicable in another (Harzing, 2006). For example, in the late 1990s, Wal-Mart entered South Korea. However, Wal-Mart Korea ultimately sold all sixteen outlets to Shinsegae, a local retailer, and left the market in 2005 (Choe, 2006). Kim (2008) argued the failure of Wal-Mart Korea came from critical shortcomings in enabling value exchange with Korean consumers, as the Korean consumers had significantly different tastes and preferences compared to American consumers. While Wal-Mart's Every-Day-Low-Price strategy fit well in North America where people are willing to compromise service and quality for price, Korean consumers were not. Koreans shop daily instead of weekly or biweekly and purchase small packages. This paper studies consumer perceptions and food groupsings in three countries: USA, China, and Korea to investigate whether there are country-specific food segments. Different strategies based on different consumers' perception for each country would be helpful for consumers and multinational companies to maximize their profits.

Given the increased focus on consumer health and well-being, it is important to consider the stability of food categorizations to changing nutrient and health information. If government policy, such as mandatory nutrient labeling, changes or if retailers adopt their own nutrient labels (such as the NuVal system or traffic light systems), prior food groupings and associations may no longer be relevant. Past research has shown that such nutritional information can alter consumer behavior (Grunert and Wills, 2007; Jo et al., 2016). Thus, this paper examines how the provision of health information influences food categorizations.

To address these issues, we conducted a study with about 600 individuals in three countries, where we solicited perceptions of the taste, healthiness, and affordability of 60 food items before and after the provision of health information. In the following section, we describe our survey and methods. Results are then discussed and the last section contains the conclusion and discussion.

Methods

Sample

We designed an online survey in Qualtrics and obtained completed responses from around 600 individuals in panels maintained by SSI in three different countries; one hundred and ninety-one people from the USA, one hundred and ninety-seven people from China, and one hundred and ninety-two people from Korea. Summary statistics describing the sample are in table 2-1. Participants ranged in age from 18 to 74 years old, and almost 50% were females in each countriy. While over half of respondents from China (68%) and Korea (67%) belong to the normal (healthy weight) category, only 35% of participants from the USA are included in the normal category, data which is consistent with national statistics on obesity prevalence. There were relatively more participants in the middle income level (between $20,000 and $80,000) in Korea (71%) compared to in the USA (51%) and China (54%). 38% of USA participants are high income category ($80,000/year or greater), which is comparatively higher than the other countries. Across the three countries, most participants in this survey are primary shoppers who are well educated and are not vegetarians.

Survey

The survey requested the participants to rate perceived taste, health, expense, and purchase intention of 60 different food items.[8] Then, the rating was repeated after the subjects had received information about each food item's healthiness. A within-subject design was constructed so that we could investigate how subjects change their perceptions according to the provision of health information and determine how this affects the food categories for CM. We randomized the order

[8] A list of 60 food items was mostly compiled based on the expenditure categories used to construct the consumer price index (CPI) released by Bureau of Labor Statistics (BLS). The CPI market basket is developed from the Consumer Expenditure Surveys for 2013 and 2014 provided from 7,000 families on what they actually bought. We used this data to identify items commonly consumed in the US. In addition, and to add diversity, we include the most expensive six foods, which are chosen according to the price, rarity, and the difficulty in the cultivating process. Since these items are not affordable for everyone, we expect them to be uncommon food items. For consistency and comparability, we applied the same list to China and Korea as well.

with 60 food items to prevent order effects. A pretest was conducted with 290 respondents to find the most efficient and accurate way to deliver health information.

In the first treatment, participants were shown a photo of each food and immediately indicated their subjective taste, health, expense perceptions and purchase intentions for each food item. Figure 2-1 is an example screen shot of the survey. For the second treatment, everything was the same as in the first treatment but it also included each item's photo and corresponding health information. The information consisted of a traffic light system (green, yellow, and red) based on the nutrient rich food (NRF) 6.3 index and energy density. The NRF 6.3 index ranks foods based on their nutrient composition. It consists of 6 qualifying nutrients—protein, fiber, iron, calcium, and vitamins A and C—and 3 disqualifying nutrients—saturated fat, added sugar, and sodium. A food's score is calculated by subtracting the sum of the percentage of the maximum recommended values for three nutrients to limit from the sum of the percentage of daily values for six nutrients to encourage.[9] The energy density represents the amount of energy per gram of food. In this study, we used the calories per 100 grams of each food item.

We conducted a cluster analysis to identify foods with similar NRF 6.3 index scores and calorie density, and we found three clusters, which we label red, yellow, and green. Foods with a green signal have positive means of NRF 6.3 index and the lowest means of energy density, while foods with a red signal have the lowest means (negative) of NRF 6.3 index and the highest means of energy density. Foods which have the highest means of NRF 6.3 index but middle level means

[9] The NR6 is calculated as $\sum_{i=1}^{6} \frac{nutrient_i}{DV_i} \times 100$ where $nutrient_i$ is *ith* nutrient per serving (g or mg) in 100g of food and DV_i is daily value for *ith* nutrient (g or mg). The LIM score is calculated as $\sum_{j=1}^{3} \frac{nutrient_j}{MRV_j} \times 100$ where $nutrient_j$ is *jth* nutrient per serving (g or mg) in 100g of food and MRV_j is maximum recommended value for *jth* nutrient (g or mg).

of energy density are located in the yellow signal. This simple type of health information should be relatively easily understood and digestible by participants.

Cluster analysis

Our data set has the average rating on perceived taste, health and expense of each food in each country. Thus, we have a total of 60 observations for each variable in each country. Based on consumers' average perceived taste, health, and expense of each food, we used k-means clustering to maximize within-group homogeneity for optimal partitions by minimizing Euclidean distances between groups[10]. Following the research of Milligan and Cooper (1985) and Calinski and Harabasz (1974), we used the pseudo F statistics to determine the appropriate number of clusters for each country.

Results

Average taste and health between countries

To understand more about consumers from the three different countries, we compare average perceived taste, health, and price and calculated Kendall's *W* statistic[11], which is a rank-based correlation measure of agreement among raters. Kendall's *W* statistic ranges from 0 to 1, where a

[10] There is no perfect consent between researchers for determining the initial seeds and the number of clusters (Everitt, 1979). Douglas (2006) synthesizes the method of initialization for the k-means clustering: randomly choosing the initial cluster seeds (McRae, 1971; Forgy, 1965; Steinley, 2003), a hybrid method combining the k-means with Ward's method (1963) (Milligan, 1980), a bootstrap-like algorithm (Bradley and Fayyard, 1998). The method of randomly choosing the initial seeds is used for this research. According to Steinley, this method outperforms several other methods.

[11] Kendall's W is defined as $w = \frac{12S}{m^2(n^3-n)}$ where S is the sum of squared deviations, $\sum_{i=1}^{n}(R_i - \bar{R})^2$, R_i is the total rank given to i th food product, $\sum_{j=1}^{m} r_{ij}$, $\bar{R}$ is the mean value of total ranks, m is the number of the country, m=1,2, and n is the number of food products, n=1,2,…,60.

0 indicates no overall agreement among countries' mean ratings and 1 indicates complete agreement.[12] Though Kendall's W is similar to correlation coefficients, the W statistic is useful in summarizing agreement when there are more than two judges (or countries). Table 2-2 shows all three countries' Kendall's W statistics for perceived taste, health, and price before and after information. Both before and after information, there are strong levels of agreement on perceived taste (0.78 for before information and 0.76 for after information), health (0.88 for before information, and 0.97 for after information), and price (0.80 for before information and 0.77 for after information) among the three countries. While the provision of health information increases the level of agreement on perceived health across countries, it does not increase the level of agreement on perceived taste and price. Interestingly, although China and Korea are within the same Asian culture area, the W statistics for perceptions are not relatively high.

The average perceived taste and health perceptions are plotted in two-dimension graphs (see Figures 2-2 to 2-7, and Appendix B). Each figure represents the average perceived taste or health of USA (or Korea) and China (or Korea) before and after the provision of information, respectively. If the foods are on the 45-degree line ($x = y$), there is perfect agreement on perceived taste (or health) about foods between the two countries. Thus, in this case, the W statistics for those products between two countries would be 1. If the foods are located on the left side of the reference line, those foods are tastier (or healthier) to consumers from the country on the y axis rather than consumers from the country on the x axis and vice versa.

Figure 2-2 shows the average perceived taste between the USA and China before consumers received health information. Orange, banana, apple, fruit juice, ice-cream, potatoes,

[12] To measure Kendall's W statistics, we ranked the average perceived taste and health (Appendix A).

chocolate, cookie, chicken, muffin, and hot dog are considered tasty foods in both countries, while margarine is considered untasty. In the graph, the circles represent the processed vegetables and fruits, either frozen or canned. The processed vegetables and fruits are tastier to Americans than to the Chinese and a similar phenomenon happens in Figure 2-3, which is for the USA and Korea. It indicates that Chinese and Korean consumers tend to consider processed vegetables and fruits less tasty than American consumers do. However, this trend changed after the provision of information in China. Figure 2-5 shows now frozen mixed vegetable, frozen mixed fruit, canned corn, and canned peach are located on the reference line. Unlike Chinese consumers, Korean consumers who received positive health information for processed vegetables and fruits still consider them less tasty than American consumers (Figure 2-6). It supports Kim's (2008) arguments that Koreans consider the freshness of food products very seriously and therefore prefer corner or wet-markets to buy small volumes of fresh products.

For healthiness, the consent across countries can be seen Appendix B. Especially for the case after consumers are provided objective health information, perceived health of food items is more densely distributed between countries compared to the plots before information. It could be seen from Table 2-2 as well. All of Kendall's W statistics for perceived health are close to 1, which means there are agreements among judges. And these statistics increase in the case of after the provision of information.

Country-specific clusters and food categories

To determine the number of clusters for the k-means model, we check the pseudo F statistics of each model from three to 60. Table 2-3 shows the results of selection statistics according to the provision of information across countries. Before respondents receive health information, the three-, three-, and six-cluster models are chosen for the USA, China, and Korea, respectively. For the case where after people are provided the information, five-, six-, and three-cluster models are selected for USA, China, and Korea, respectively. The provision of information changed the cluster model in all three countries. While the number of food segments for the USA and China increased, Korea's number of clusters decreased as people received the nutrient information. One possible explanation is that the provision of information causes Korean consumers to have similar taste, health, and price perceptions, and, consequently leading to the smaller number of optimal partitions which maximize within-group homogeneity.

Appendix C shows the results of cross-countries' k-mean cluster analysis according to the provision of information, and Tables 2-4 to 2-6 indicate the mean values of consumers' perceived taste, healthiness, expense, and purchase intention for each cluster on a scale from
-5 to 5. Before information for the USA, the food items fall into three clusters which we call *Hedonic, Uncommon,* and *Ideal* food clusters. Twenty-one food items are included in the *Hedonic* cluster, and the average taste is the highest among all clusters while the average healthiness is lowest. Also, these food items are the most affordable foods. Unlike the *Hedonic* cluster, the *Ideal* cluster consists of food items which are the healthiest. People would like to purchase foods from *Ideal* cluster the most. Lastly, Beluga caviar, Foie gras, White truffle, Saffron, Donkey cheese, and Frozen scallop are included in the *Uncommon* cluster, which are perceived as the least tasty, and the most expensive. Consumers are least likely to purchase *Uncommon* cluster foods compared with the other two cluster foods.

After the provision of information, people changed their perceived taste, health, expense, and purchase intention of 60 food items, and it leads now to the five-cluster model—*Taste-oriented, Ideal, Uncommon, Moderately Ideal, Health-oriented.* As consumers receive objective health signal information, instead of using the *Hedonic* cluster which is the highest in average taste, the lowest in average health, and the most affordable, *Taste-oriented*, *Moderately Ideal*, and *Health-oriented* clusters are generated. However, *Uncommon* and *Ideal* clusters still remained after the provision of information.

Beluga caviar, Foie gras, White truffle, Saffron, and Donkey cheese are in the *Uncommon* cluster and consumers consider them untasty, unhealthy, and expensive. This result supports the finding of Quealy and Sanger-Katz (2016), who conducted a survey to a panel of nutrient experts and Americans about which foods they thought were good or bad for you. They found that nutritionists' healthiness ratings for quinoa, tofu, sushi, and hummus are higher than those of the public. Being that many of them are new foods in the mainstream American diet, they concluded that Americans tend to consider foods that are unfamiliar as not healthy. All foods from the *Ideal* cluster are originally healthy foods according to either nutrient or energy density. When we consider that the *Ideal* cluster contains eight yellow signals and one red signal under the case of before information, changes in the *Ideal* cluster could provide the evidence of information updates. Further, this cluster consists of the most tasty, the most healthy, the most affordable, and the most likely to be purchased foods. As the second most highly preferred food group, the *Moderately Ideal* cluster contains relatively tasty and healthy foods. The *Taste-oriented* cluster consists of relatively tasty, the least healthy, and the most affordable food items. On the other hand, the foods in which average values of health are higher than that of taste are included in the *Health-oriented* cluster. Comparison between these two clusters indicates that people tend to have higher purchase intensions for *Taste-oriented* foods than for *Health-oriented* foods.

China has the three-cluster model before consumers receive the health information. Though it has three clusters like the USA model, the propensity of clusters is different. Instead of the *Hedonic* and *Uncommon* clusters of the USA model, *Health-oriented* and *Taste-oriented* clusters are generated. *Health-oriented* products are more often considered expensive but, interestingly, more likely to be purchased compared to the *Taste-oriented* cluster. Since foods are necessary products, they are supposed to have a negative relationship between price and choice. However, it would not matter since the average expense of all three clusters has negative values, which means people already think the price of products is affordable enough.

After the provision of health information, the cluster model changed from the three-cluster model to the six-cluster model—*Ideal, Uncommon, Less taste oriented, Unfavorable, Taste oriented,* and *Moderately Ideal*. That is, health information makes consumers' perceptions more sparsely distributed. Overall, the average expense is negative across clusters, which means consumers consider all products affordable enough. Intriguingly, in China, the correlation coefficient between perceived taste and health increased from 0.59 to 0.83 with the provision of objective health information, which means now consumers tend to consider tasty (or healthy) foods are healthy (or tasty). This correlation coefficient is high relative to that of USA and Korea, which are 0.12 to 0.36 and -0.02 to 0.28, respectively. Thus, China's cluster model does not have the *Hedonic* cluster, which is the highest in taste and the lowest in health. Also, foods with the highest in average perceived taste and the highest in average perceived health are in the *Ideal* cluster, and foods with the second highest in average perceived taste and the second highest in average perceived are included in the *Moderately Ideal* cluster. Another fascinating point is that the *Uncommon* cluster is generated after the information is provided. While products of the *Uncommon* cluster in the USA are not only untasty but also unhealthy and expensive, products of the *Uncommon* cluster in China are considered untasty, but relatively healthy and the most expensive.

Korea has the six-cluster model before people receive health information: *Less taste-oriented, Less health-oriented, Ideal, Hedonic, Taste-oriented,* and *Health-oriented* clusters. Unlike in the USA and China, the provision of health information has a different influence on the cluster model of Korea. Consumers in Korea tend to have a certain agreement of perception and it leads to a decreased number of clusters after information from six to three. The three-cluster model contains *Health-oriented*, *Ideal*, and *Taste-oriented*. Consumers consider *Health-oriented* products more expensive and more likely to be purchased than *Taste-oriented* products. However, before the information, they were willing to purchase *Hedonic* products rather than *Health-oriented* (or *Less health-oriented*) products. This would be a good example of enhancing consumers' healthy diet and the nutrient-to-energy ratio.

Strategies for suppliers and retailers by clusters

The multi attribute utility theory (MAUT) is the model for describing the preferences of the decision maker over a subset of objectives (Keeney and Raiffa, 1976). MAUT assumes that decision makers express their preferences based on multiple attributes, and either explicitly weigh the alternatives or make mental representations of choices before deciding what actions to take (Glanz et al., 1998). Thus, based on MAUT assumption, we estimated linear regression models for each cluster to investigate how consumers' perceived taste, health, and expense affect their purchase intentions. In all clusters, perceived taste and health have a positive relationship with purchase intention. This result has a thread of connections with previous literature, saying taste and health are the most important two factors when consumers purchase. Also, it provides the basis for why suppliers and retailers should produce products that look more tasty and healthy to attract consumers' interest. Advertisements emphasizing tastiness and healthiness of products, or functional foods which (a food given a health-promotion or disease prevention), would help to increase their sales and market share.

The regression results for most clusters show a negative sign effect of expense, which is consistent with demand theory. Since consumers are willing to purchase more if the price is expected to be more affordable, some low price strategy—price promotion, store brand, and so on—could increase the profit of suppliers and retailers. However, ten clusters have a price coefficient which is not significant at the 5% level: the *Hedonic* cluster (USA, before information), *Ideal* cluster (USA, before information), *Taste oriented* cluster (USA, after information), *Taste oriented* cluster (China, before information), *Unfavorable* cluster (China, after information), *Less taste oriented* cluster (Korea, before information), *Ideal* cluster (Korea, before information), *Taste oriented* cluster (Korea, before information), *Ideal* cluster (Korea, after information), and *Taste oriented* cluster (Korea, after information). Since the price of these foods would not significantly influence consumers' purchase intentions, suppliers and retailers do not need to pursue a low price policy to increase their sales.

Also there are three clusters which have a positive relationship between perceived expense and purchase intention: *Ideal* cluster (USA, after information), *Health-oriented* cluster (USA, after information), and *Hedonic* cluster (Korea, before information). The positive relationship indicates that consumers tend to purchase more food items if they are perceived as expensive. In the USA, this phenomenon is observed in *Ideal* and *Health-oriented* clusters after consumers receive health information, which implies more expensive prices could be a signal of healthier foods in the situation where consumers could have objective health information. On the other hand, in the case of Korea, a positive relationship is found in the *Hedonic* cluster before the provision of information. In other words, when Koreans do not have objective health information, they are more willing to purchase expensive *Hedonic* cluster foods than affordable *Hedonic* cluster foods, which are bacon, sausage, ice-cream, doughnut, pizza, and hamburger. Surprisingly, in the USA and China, consumers' purchase intentions for these kinds of foods are not affected by the price.

Conclusion and Discussion

In this study, we create consumer-oriented food clusters using cluster analysis. These food clusters may be useful for CM strategies. The resulting food clusters do not necessarily indicate which products should be situated close to each other in a retail establishment; but they do provide potential groupings of similar foods in the consumers' minds. Foods in a common cluster are likely to be relatively substitutable, and as such it might be possible to use these results to decrease inventory management costs or to select items to be included in a store. For instance, foods in the Uncommon cluster are considered the most expensive, least tasty, and least preferred to purchase by American and Chinese consumers. Thus, these products are not necessarily included on store shelves in the USA and China to increase retailer benefits.

In the USA, price could be a signal about healthy foods in certain categories. Americans are more willing to purchase expensive healthy foods rather than affordable healthy foods if they have objective health information. For the *Hedonic* or *Taste-oriented* products, such as bacon, hamburger, candy, and butter, price does not imply additional information and it would not affect consumers' purchase intentions in the USA. Thus, both low price promotion and luxury brand strategy will not be very effective. For China, consumers tend to consider healthy foods tasty as well after the provision of information. Thus, to improve sales, advertising which emphasizes healthiness of products would be effective. Further, for most foods, consumers in China would like to purchase more for affordable products rather than expensive products.

In the situation where Korean consumers do not have objective health information, for Hedonic cluster products—bacon, sausage, ice cream, doughnut, and so on—they are willing to purchase more expensive ones rather than relatively affordable ones. Thus, a luxury brand strategy would be more effective to increase sales than low price promotion. However, in the case where products are provided with health information, focusing on taste or improving healthiness would be more helpful to maximize profits rather than price strategy. Also, concerning *Ideal* cluster products—apple, banana, chicken, salad, and so on—price would not affect purchase intentions in both with and without the provision of information.

Identifying consumer-oriented food clusters would be helpful for efficient category reduction and improving healthy dietary patterns. Retailers and suppliers could use food classifications to implement appropriate strategies by each cluster to increase margin and market shares. Multinational companies could also use food clusters for efficient localization. One limitation of this study is that it does not provide a within-products level categoriations—e.g., Fuji apple, jazz apple, and gala apple. In grocery retail setting, a lower level categorization might be useful to organize shelves at the store. This study provided a first step in attempting to understand how consumers in three different countries classify diverse foodstuffs. Future research will be needed to explore how such categorizations can help increase profitably for retailers.

References

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, *3*(1), 1-27.

Choe, S. H. (2006, May 23). Wal-Mart Selling Stores and Leaving South Korea. *The New York Times*, Retrieved from http://www.nytimes.com

Desrochers, D.M., & Nelson, P. (2006). Adding consumer behavior insights to category management: Improving item placement decisions. *Journal of Retailing*, *82*(4), 357-365.

Drewnowski, A. (2010). The Nutrient Rich Foods Index helps to identify healthy, affordable foods. *American Journal of Clinical Nutrition*, *91*(4), 1095S-1101S.

Drewnowski, A., & Fulgoni, V. (2008). Nutrient profiling of foods: creating a nutrient-rich food index. *Nutrition Reviews*, *66*(1), 23-39.

Drewnowski, A. (2005). Concept of a nutritious food: toward a nutrient density score. *American journal of clinical nutrition*, *82*(4), 721-732.

Dudlicek, J. (2016, June 29). Nielsen Consumer 360: Knowing the Why Behind Shoppers. *Progressive Grocer,* Retrieved from http://www.progressivegrocer.com

Glanz, K., Basil, M., Maibach, E., Goldberg, J., & Snyder, D. A. N. (1998). Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *Journal of the American Dietetic Association*, *98*(10), 1118-1126.

Grant, A. W., & Schlesinger, L. A. (1995). Realize your customers' full profit potential. *Harvard Business Review*, *73*(5), 59-72.

Grunert, K. G., & Wills, J. M. (2007). A review of European research on consumer response to nutrition information on food labels. *Journal of Public Health*, *15*(5), 385-399.

Harzing, A. W. (2006). Response styles in cross-national survey research a 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243-266.

Holweg, C., Schnedlitz, P., & Teller, C. (2009). The Drivers of Consumer Value in the ECR Category Management Model. *The International Review of Retail, Distribution and Consumer Research, 19*(3), 199-218.

Jo, J., Lusk, J.L., Muller, L., & Ruffieux, B. (2016). Value of parsimonious nutritional information in a framed field experiment. *Food Policy*, *63*, 124-133.

Joint Industry Project on Efficient Consumer Response. (1995). *Category Management Report: Enhancing Consumer Value in the Grocery Industry.*

Karolefski, J. (2016, February 10). Category Management is Due for Radical Change. *Progressive Grocer,* Retrieved from http://www.progressivegrocer.com

Keeney, & Raiffa, H. (1976). *Decisions with multiple objectives*. New York: John Wiley.

Kim, R. B. (2008). Wal-Mart Korea: Challenges of entering a foreign market. *Journal of Asia-Pacific Business*, *9*(4), 344-357.

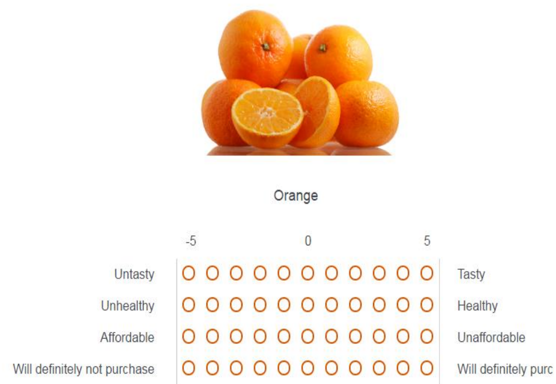
Lusk, J. L., & Briggeman, B. C. (2009). Food values. *American Journal of Agricultural Economics*, *91*(1), 184-196.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159-179.

Quealy, K., & Sanger-Katz, M. (2016, July 5). Is Sushi 'Healthy'? What About Granola? Where Americans and Nutritionists Disagree. *The New York Times,* Retrieved from http://www.nytimes.com

Zakowska-Biemans, S. (2011). Polish consumer food choices and beliefs about organic food. *British Food Journal*, *113*(1), 122-137.

First treatment (Before information)	Second treatment (After information)



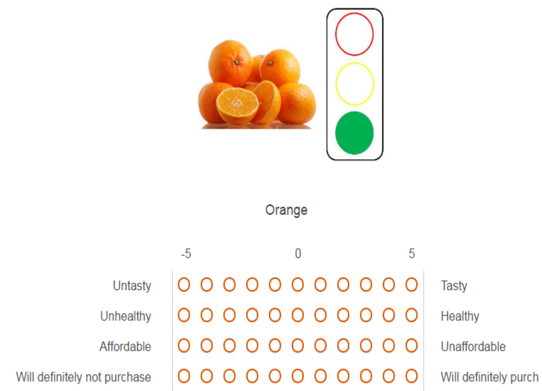


Figure 2-1. Screen shot of the survey

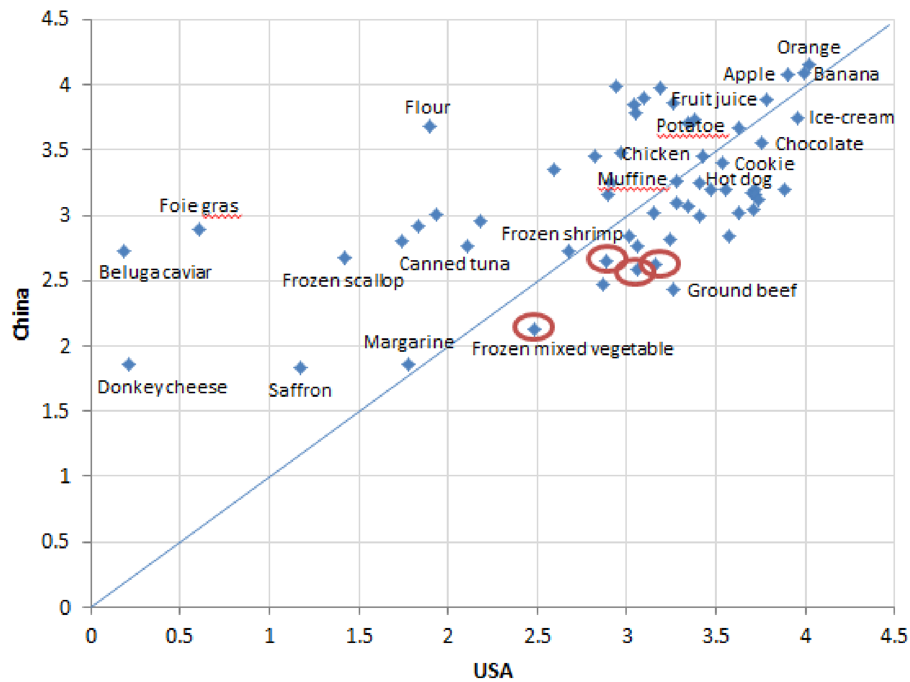


Figure 2-2. Average perceived taste in USA and China before information (Note: The red circles represent the processed vegetables and fruits; frozen mixed vegetables, frozen mixed fruits, canned corn, and canned peach.)

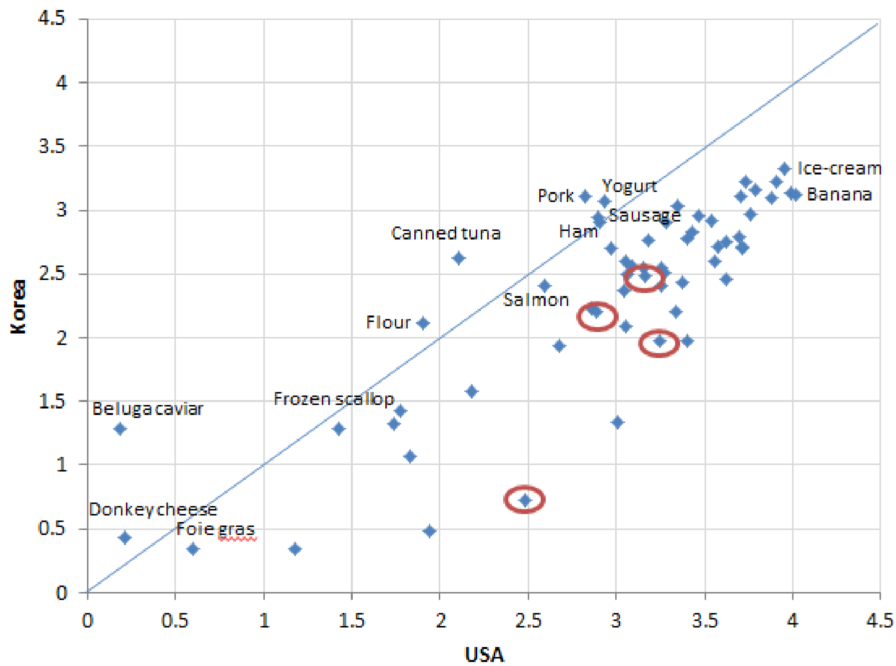


Figure 2-3. Average perceived taste in USA and Korea before information (Note: The red circles represent the processed vegetables and fruits; frozen mixed vegetables, frozen mixed fruits, canned corn, and canned peach.)

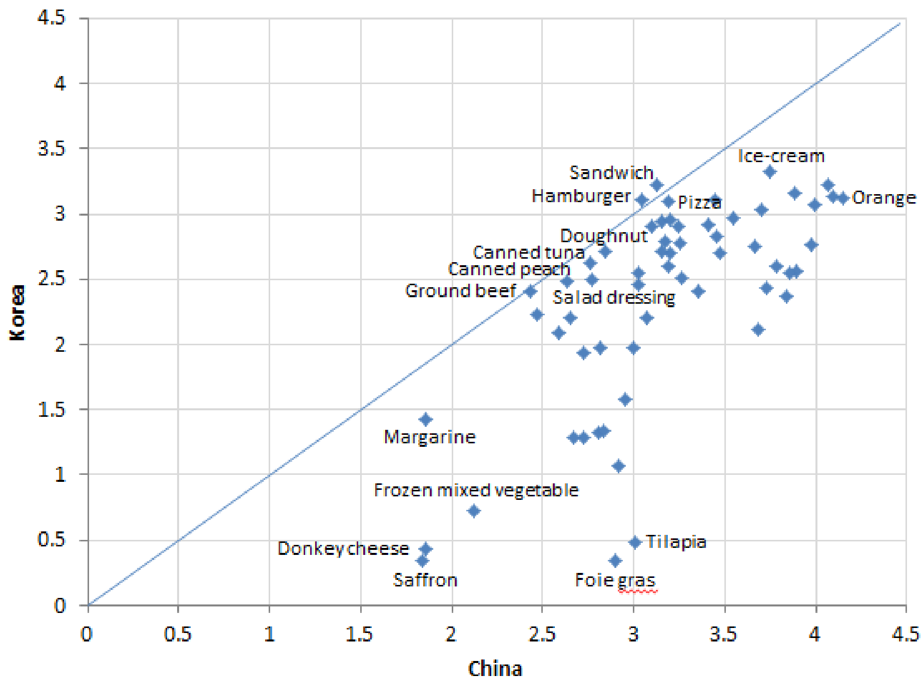


Figure 2-4. Average perceived taste in China and Korea before information

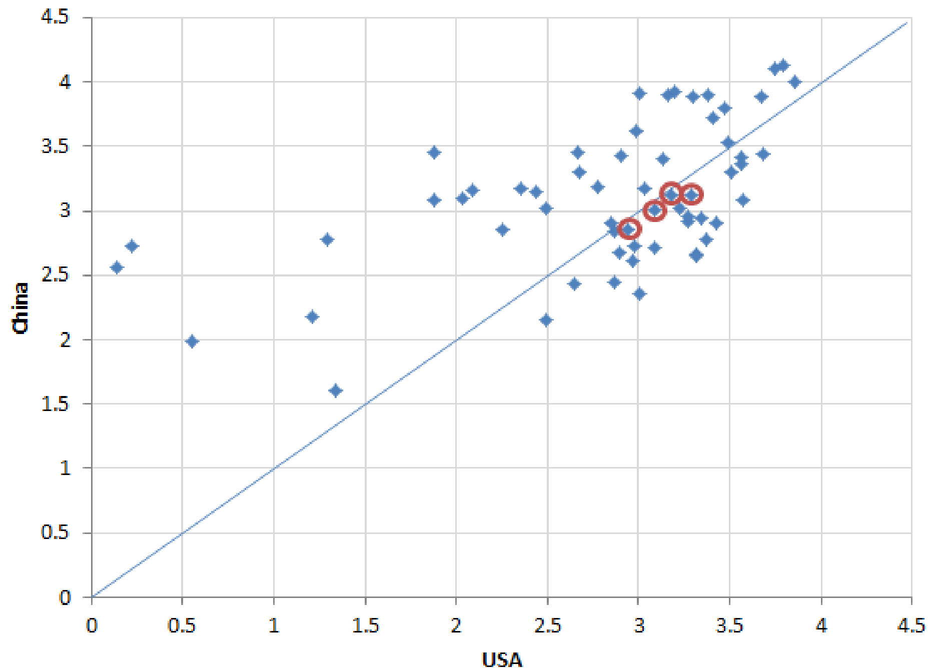


Figure 2-5. Average perceived taste in USA and China after information (Note: The red circles represent the processed vegetables and fruits; frozen mixed vegetables, frozen mixed fruits, canned corn, and canned peach.)

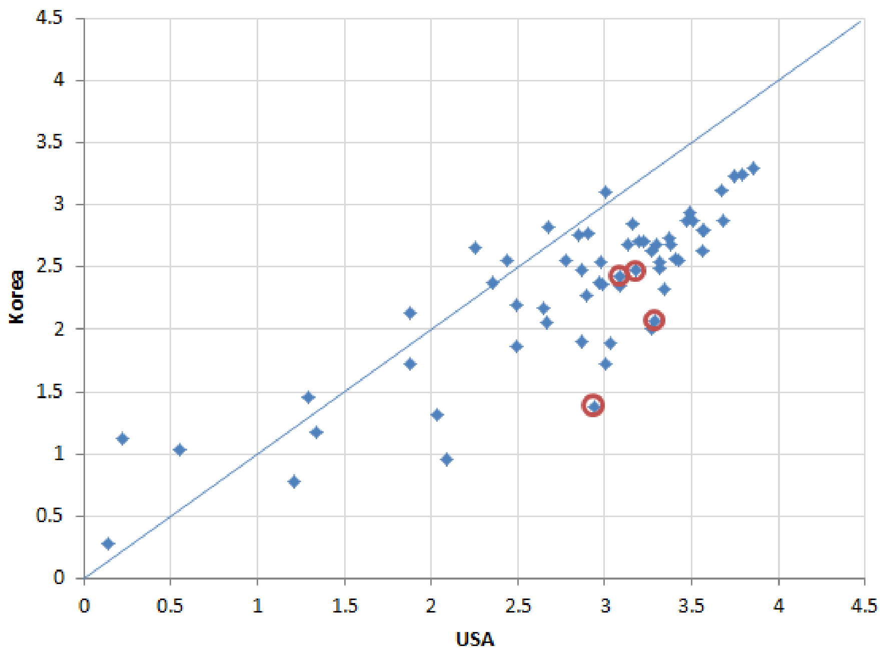


Figure 2-6. Average perceived taste between USA and Korea after information (Note: The red circles represent the processed vegetables and fruits; frozen mixed vegetables, frozen mixed fruits, canned corn, and canned peach.)

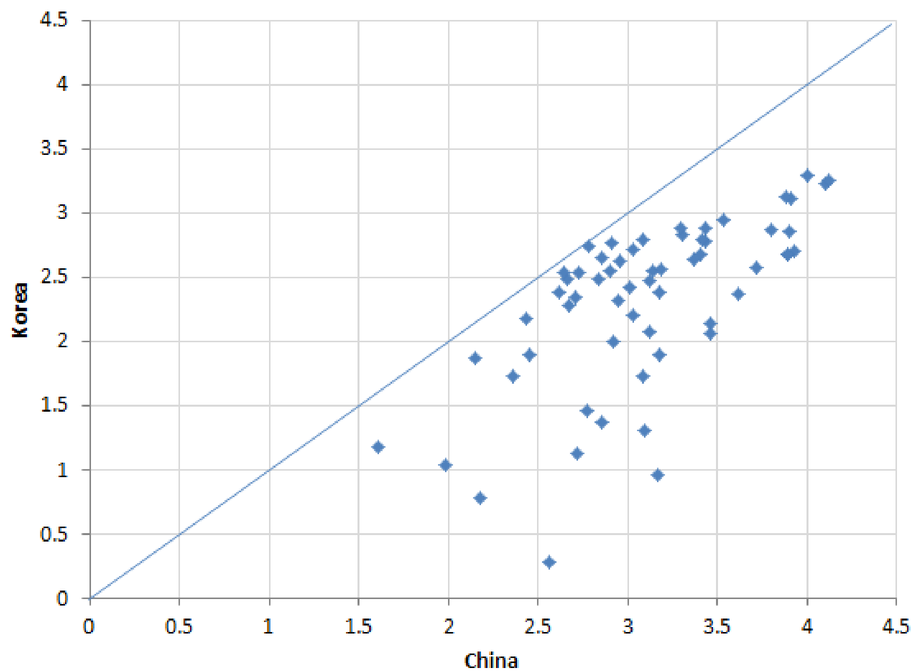


Figure 2-7. Average perceived taste between China and Korea after information

Table 2-1. Socio-demographic characteristics of the sample (%)

Characteristics	Category	USA	China	Korea
Total	n	191	197	192
Age	18 - 24 years old	15	13	7
	25 - 34 years old	39	41	22
	35 - 44 years old	25	34	35
	45 - 54 years old	9	12	27
	55 - 64 years old	12	1	9
	65 - 74 years old	1	0	0
Gender	Female	49	55	45
BMI	Underweight	7	8	7
	Normal(Healthy weight)	35	68	67
	Overweight	28	21	23
	Obese	30	4	3
Income	Low level (< $20,000/year)	11	19	14
	Middle level ($20,000 - $80,000/year)	51	54	71
	High level(> $80,000/year)	38	27	15
Primary Shopper	Primary shopper	82	81	67
	Co-shopper	14	10	19
	None	4	10	14
Vegetarian or Vegan	Vegetarian or Vegan	11	19	7
Education	> BA/BS college degree	51	70	64

Table 2-2. Kendall's W statistics of perceived taste and health among three countries

	Country	Taste	Health	Price
Before Information	USA, China, and Korea	0.78	0.88	0.80
	USA and China	0.77	0.90	0.83
	USA and Korea	0.88	0.92	0.92
	China and Korea	0.85	0.92	0.79
After Information	USA, China, and Korea	0.76	0.97	0.77
	USA and China	0.76	0.98	0.75
	USA and Korea	0.87	0.97	0.94
	China and Korea	0.84	0.98	0.78

Note: Kendall's *W* statistic ranges from 0 to 1, where a 0 indicates no overall agreement among countries' mean ratings and 1 indicates complete agreement

Table 2-3. Selection statistic for determining number of clusters (k)

	Before Information (Pseudo F Statistic)			After Information (Pseudo F Statistic)		
k	USA	China	Korea	USA	China	Korea
9	-	-	42.1	-	-	-
8	-	-	42.7	-	-	-
7	-	60.9	43.0	-	-	-
6	46.3	62.7	**48.0**	88.1	**145.8**	-
5	49.0	61.5	39.9	**92.4**	135.8	49.0
4	48.5	70.0	41.1	77.5	99.9	41.4
3	**57.0**	**75.4**	44.4	67.0	138.4	**60.5**

Note: Bold indicates the largest values of Pseudo F statistic and k which matches with each bold is selected for the number of clusters of k-means process.

Table 2-4. Average perceived taste, health, expense, and purchase intention cross clusters for USA

	Cluster	Num of Foods	Taste	Health	Expense	Purchase Intention
Before Information	Ideal	33	3.15	2.85	-0.92	2.61
	Hedonic	21	3.22	0.30	-1.01	2.26
	Uncommon	6	0.89	1.07	0.82	-0.20
After Information	Ideal	21	3.30	3.31	-1.09	2.85
	Moderately Ideal	16	3.10	1.10	-1.08	2.51
	Taste oriented	12	2.78	-0.92	-1.39	1.62
	Health oriented	6	2.29	2.54	-0.07	1.51
	Uncommon	5	0.68	0.26	0.52	-0.43

Table 2-5. Average perceived taste, health, expense, and purchase intention cross clusters for China

	Cluster	Num of Foods	Taste	Health	Expense	Purchase Intention
Before Information	Ideal	19	3.70	3.73	-2.60	3.45
	Health oriented	31	2.89	2.26	-1.85	1.92
	Taste oriented	10	2.93	0.76	-2.36	1.78
After Information	Ideal	12	3.88	3.96	-2.82	3.75
	Moderately Ideal	18	3.25	2.98	-2.40	2.70
	Taste oriented	12	2.97	1.58	-2.35	2.13
	Less taste oriented	10	2.67	-0.05	-2.35	1.34
	Uncommon	5	2.57	1.80	-1.43	1.23
	Unfavorable	3	2.07	-0.60	-2.36	0.59

Table 2-6. Average perceived taste, health, expense, and purchase intention cross clusters for Korea

	Cluster	Num of Foods	Taste	Health	Expense	Purchase Intention
Before Information	Ideal	12	2.86	2.84	0.94	2.42
	Taste oriented	14	2.52	1.07	1.18	1.55
	Hedonic	15	2.79	-0.43	1.12	1.29
	Health oriented	6	2.07	2.41	2.61	0.95
	Less taste oriented	5	2.11	-0.85	-0.38	0.65
	Less health oriented	8	0.75	1.17	1.56	-0.19
After Information	Ideal	13	2.90	3.01	1.06	2.44
	Health oriented	29	2.12	1.29	1.45	1.12
	Taste oriented	18	2.22	-0.86	0.79	0.62

Appendix A.

Table A2-1 Rank of average perceived taste

	Before information			After information		
Rank	USA	China	Korea	USA	China	Korea
1	Orange	Orange	Ice cream	Banana	Apple	Banana
2	Banana	Banana	Sandwich	Apple	Orange	Apple
3	Ice cream	Apple	Apple	Orange	Banana	Orange
4	Apple	Yogurt	Fruit juice	Salad	Milk	Fruit juice
5	Pizza	Yubari	Banana	Fruit juice	Yogurt	Yogurt
6	Fruit juice	Rice	Orange	Pizza	Yubari	Meat-chicken
7	Chocolate	Fruit juice	Meat-pork	Sandwich	Tomato	Salad
8	Sandwich	Tomato	Hamburger	Burrito	Lettuce	Ice cream
9	Cheese	Lettuce	Pizza	Ice cream	Fruit juice	Potato
10	French fries	Milk	Yogurt	Meat-chicken	Potato	Yubari
11	Hamburger	Ice cream	Meat-beef	Potato	Soup	Meat-pork
12	Salad	Soup	Chocolate	French fries	Rice	Sandwich
13	Cereal	Meat-beef	Chicken tender	Soup	Meat-chicken	Pizza
14	Potato	Flour	Sausage	Tomato	Vegetable juice	Ham
15	Doughnut	Potato	Cookie	Hamburger	Flour	Hot dog
16	Burrito	Chocolate	Ham	Cereal	Salad	Hamburger
17	Cookie	Roasted beef	Hot dog	Cookie	Ham	Chicken tender
18	Chicken tender	Meat-chicken	Meat-chicken	Cheese	Sandwich	Milk
19	Meat-chicken	Meat-pork	Salad	Lettuce	Meat-beef	Tomatoe
20	Bacon	Cookie	Bacon	Frozen mixed fruit	Burrito	Lettuce
21	Peanut butter	Salmon	Yubari	Chocolate	Meat-pork	Meat-beef
22	Soup	Muffin	Potato	Pasta	Ice cream	Sausage
23	Meat-beef	Bacon	French fires	Chicken tender	Roasted beef	Burrito
24	Pasta	Ham	Doughnut	Milk	Salmon	Chocolate

25	Hot dog	Cheese	Roasted beef	Canned peach	Meat-turkey	Soup
26	Muffin	Chicken tender	Cheese	Yubari	Tilapia	Roasted beef
27	Ground beef	Pizza	Canned tuna	Meat-beef	Canned tuna	Canned tuna
28	Tomato	Burrito	Milk	Ground beef	Frozen mixed fruit	French fries
29	Frozen mixed fruit	Salad	Burrito	Canned corn	Canned peach	Doughnut
30	Yubari	French fries	Rice	Meat-turkey	Catfish	Cookie
31	Canned peach	Sausage	Tomato	Yogurt	Frozen scallop	Cheese
32	Sandwich bread	Sandwich	Sandwich bread	Peanut butter	Pizza	Sandwich bread
33	Rice	Hot dog	Muffin	Rice	Chicken tender	Canned peach
34	Salad dressing	Pasta	Salad dressing	Doughnut	Frozen shrimp	Canned corn
35	Candy	Hamburger	Canned peach	Bacon	Canned corn	Salmon
36	Milk	Cereal	Cereal	Frozen mixed vegetables	Chocolate	Bacon
37	Lettuce	Sandwich bread	Soup	Ham	Cereal	Rice
38	Meat-turkey	Tilapia	Salmon	Muffin	Pasta	Ground beef
39	Roasted beef	Peanut butter	Ground beef	Candy	Hot dog	Cereal
40	Yogurt	Vegetable juice	Lettuce	Sandwich bread	French fries	Muffin
41	Ham	Catfish	Butter	Hot dog	Frozen mixed vegetables	Frozen shrimp
42	Sausage	Foie gras	Pasta	Roasted beef	Sausage	Salad dressing
43	Canned corn	Doughnut	Canned corn	Meat-pork	Sandwich bread	Flour
44	Butter	Meat-turkey	Flour	Vegetable juice	Hamburger	Frozen mixed fruit
45	Meat-pork	Frozen mixed fruit	Candy	Salad dressing	White truffle	Vegetable juice

46	Frozen shrimp	White truffle	Peanut butter	Butter	Doughnut	Pasta
47	Salmon	Salad dressing	Frozen mixed fruit	Frozen shrimp	Beluga caviar	Candy
48	Frozen mixed vegetables	Canned tuna	Frozen shrimp	Canned tuna	Ground beef	Meat-turkey
49	Vegetable juice	Frozen shrimp	Vegetable juice	Salmon	Muffin	Butter
50	Canned tuna	Beluga caviar	Margarine	Sausage	Cheese	Frozen scallop
51	Tilapia	Frozen scallop	Meat-turkey	Tilapia	Cookie	Peanut butter
52	Flour	Canned corn	White truffle	Catfish	Bacon	White truffle
53	Catfish	Canned peach	Frozen scallop	Flour	Foie gras	Frozen mixed vegetables
54	Margarine	Candy	Beluga caviar	Frozen scallop	Candy	Catfish
55	White truffle	Butter	Catfish	Margarine	Salad dressing	Margarine
56	Frozen scallop	Ground beef	Frozen mixed vegetables	White truffle	Peanut butter	Beluga caviar
57	Saffron	Frozen mixed vegetables	Tilapia	Saffron	Saffron	Donkey cheese
58	Foie gras	Margarine	Donkey cheese	Donkey cheese	Butter	Tilapia
59	Donkey cheese	Donkey cheese	Foie gras	Beluga caviar	Donkey cheese	Saffron
60	Beluga caviar	Saffron	Saffron	Foie gras	Margarine	Foie gras

Table A2-2 Rank of average perceived health

Before information				After information		
Rank	USA	China	Korea	USA	China	Korea
1	Apple	Apple	Tomato	Apple	Apple	Apple
2	Banana	Lettuce	Apple	Orange	Orange	Tomato
3	Orange	Tomato	Vegetable juice	Banana	Banana	Lettuce
4	Lettuce	Banana	Lettuce	Lettuce	Lettuce	Vegetable juice
5	Tomato	Orange	Orange	Salad	Tomato	Banana
6	Vegetable juice	Milk	Banana	Tomato	Milk	Milk
7	Salad	Yogurt	Milk	Yubari	Potato	Orange
8	Yubari	Rice	Yogurt	Vegetable juice	Fruit juice	Yogurt
9	Salmon	Yubari	Potato	Frozen mixed vegetables	Yogurt	Potato
10	Yogurt	Flour	Salad	Fruit juice	Yubari	Yubari
11	Meat-chicken	Potato	White truffle	Meat-chicken	Vegetable juice	Fruit juice
12	Milk	Soup	Yubari	Milk	Soup	Salad
13	Fruit juice	Vegetable juice	Salmon	Soup	Meat-chicken	Meat-chicken
14	Frozen mixed vegetables	Fruit juice	Cheese	Yogurt	Salad	Frozen mixed fruit
15	Meat-turkey	Meat-beef	Fruit juice	Frozen mixed fruit	Catfish	Soup
16	Frozen mixed fruit	Salmon	Meat-chicken	Meat-turkey	Meat-turkey	Meat-turkey
17	Soup	Cereal	Meat-beef	Potato	Frozen mixed vegetables	Catfish
18	Potato	Salad	Catfish	Canned tuna	Tilapia	Frozen mixed vegetables
19	Rice	Catfish	Beluga caviar	Canned peach	Frozen mixed fruit	Frozen scallop
20	Tilapia	Meat-chicken	Meat-pork	Canned corn	Burrito	Salmon
21	Sandwich bread	White truffle	Soup	Sandwich	Frozen shrimp	Canned tuna

22	Catfish	Tilapia	Meat-turkey	Tilapia	Frozen scallop	Frozen shrimp
23	Frozen shrimp	Meat-pork	Canned tuna	Frozen shrimp	Rice	Cheese
24	Canned tuna	Pasta	Frozen mixed fruit	Catfish	Canned corn	Tilapia
25	Cheese	Beluga caviar	Ground beef	Frozen scallop	Canned tuna	White truffle
26	Canned peach	Sandwich bread	Saffron	Salmon	Sandwich	Meat-beef
27	Sandwich	Meat-turkey	Rice	Ham	Canned peach	Burrito
28	Canned corn	Saffron	Donkey cheese	Burrito	Flour	Sandwich
29	Peanut butter	Roasted beef	Frozen scallop	Rice	Ham	Meat-pork
30	Cereal	Cheese	Frozen shrimp	Sandwich bread	Meat-beef	Canned corn
31	Meat-beef	Frozen mixed fruit	Salad dressing	Cereal	Salmon	Ground beef
32	Pasta	Foie gras	Tilapia	Meat-beef	Cereal	Rice
33	Frozen scallop	Peanut butter	Frozen mixed vegetables	Pasta	Meat-pork	Canned peach
34	Roasted beef	Ground beef	Roasted beef	Cheese	White truffle	Beluga caviar
35	Saffron	Burrito	Cereal	Ground beef	Pasta	Saffron
36	Ground beef	Muffin	Sandwich	Roasted beef	Saffron	Donkey cheese
37	Meat-pork	Bacon	Foie gras	Flour	Beluga caviar	Roasted beef
38	Flour	Frozen shrimp	Sandwich bread	Meat-pork	Roasted beef	Ham
39	Ham	Cookie	Burrito	Saffron	Sandwich bread	Cereal
40	White truffle	Frozen scallop	Pasta	Chicken tender	Ground beef	Sandwich bread
41	Salad dressing	Frozen mixed vegetables	Canned corn	Pizza	Cheese	Pasta
42	Chocolate	Salad dressing	Chocolate	White truffle	Chicken tender	Chicken tender
43	Beluga caviar	Canned tuna	Bacon	Beluga caviar	Pizza	Flour

44	Chicken tender	Pizza	Butter	Hamburger	Donkey cheese	Hot dog
45	Burrito	Donkey cheese	Chicken tender	Donkey cheese	Ice cream	Pizza
46	Muffin	Canned corn	Canned peach	Ice cream	Hot dog	Ice cream
47	Foie gras	Butter	Muffin	Peanut butter	Hamburger	Salad dressing
48	Butter	Sandwich	Flour	French fries	Peanut butter	Hamburger
49	Pizza	Chocolate	Ham	Hot dog	French fries	Foie gras
50	Donkey cheese	Ham	Sausage	Chocolate	Foie gras	Bacon
51	Sausage	Chicken tender	Peanut butter	Salad dressing	Salad dressing	French fries
52	Bacon	Canned peach	Cookie	Foie gras	Chocolate	Butter
53	Ice cream	Sausage	Hot dog	Muffin	Cookie	Chocolate
54	Hamburger	Hot dog	Ice cream	Butter	Muffin	Muffin
55	Margarine	Doughnut	Pizza	Bacon	Bacon	Sausage
56	Cookie	Ice cream	Margarine	Cookie	Sausage	Cookie
57	French fries	Hamburger	Doughnut	Sausage	Doughnut	Peanut butter
58	Hot dog	Candy	Hamburger	Margarine	Butter	Margarine
59	Doughnut	Margarine	French fries	Doughnut	Candy	Doughnut
60	Candy	French fries	Candy	Candy	Margarine	Candy

Appendix B

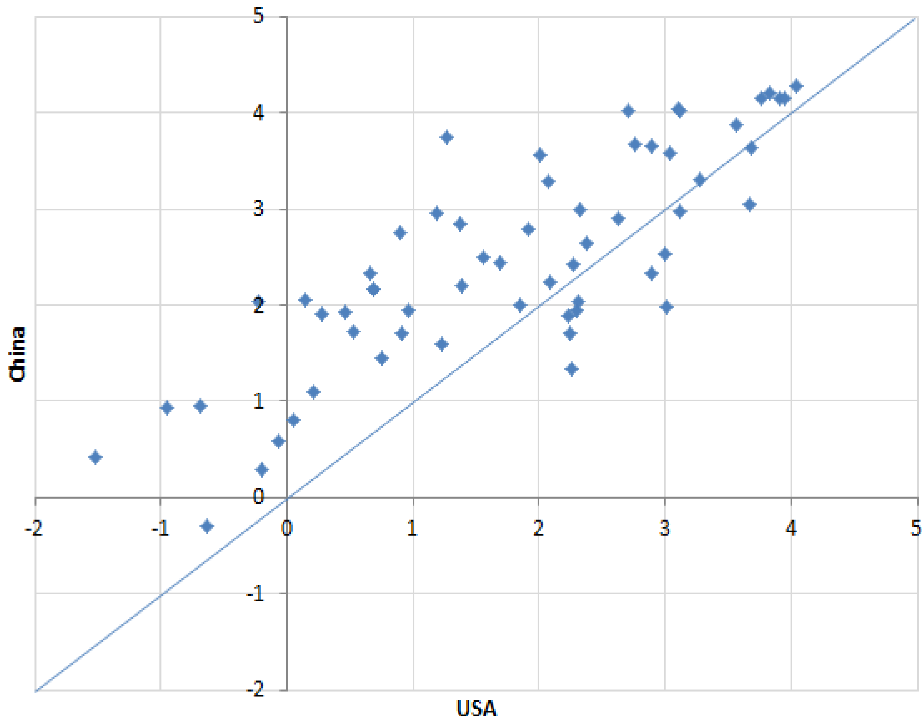


Figure B2-1 Average perceived health in USA and China before information

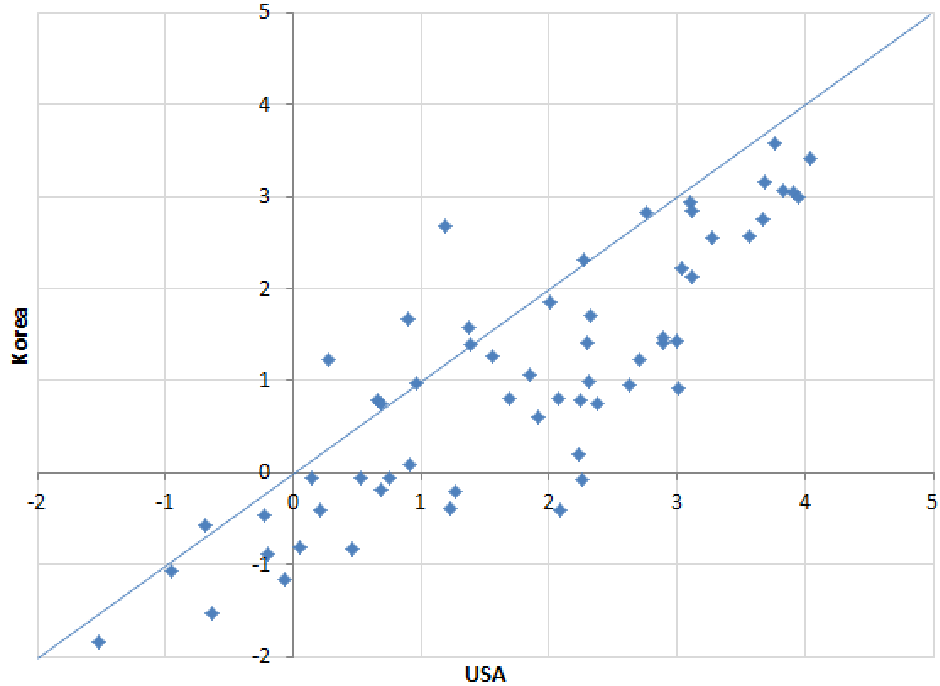


Figure B2-2 Average perceived health in USA and Korea before information

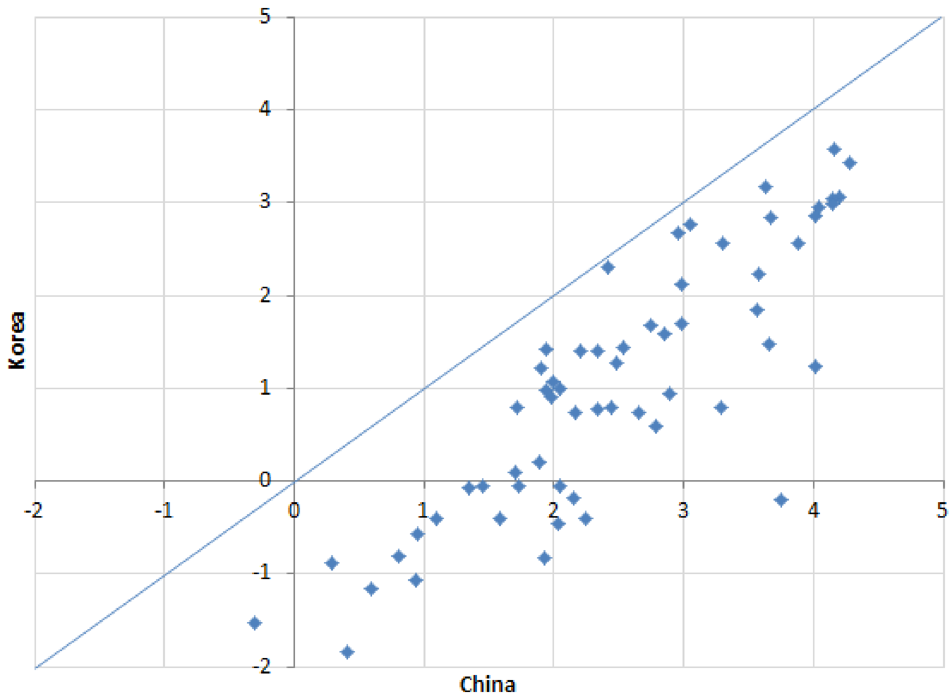


Figure B2-3 Average perceived health in China and Korea before information

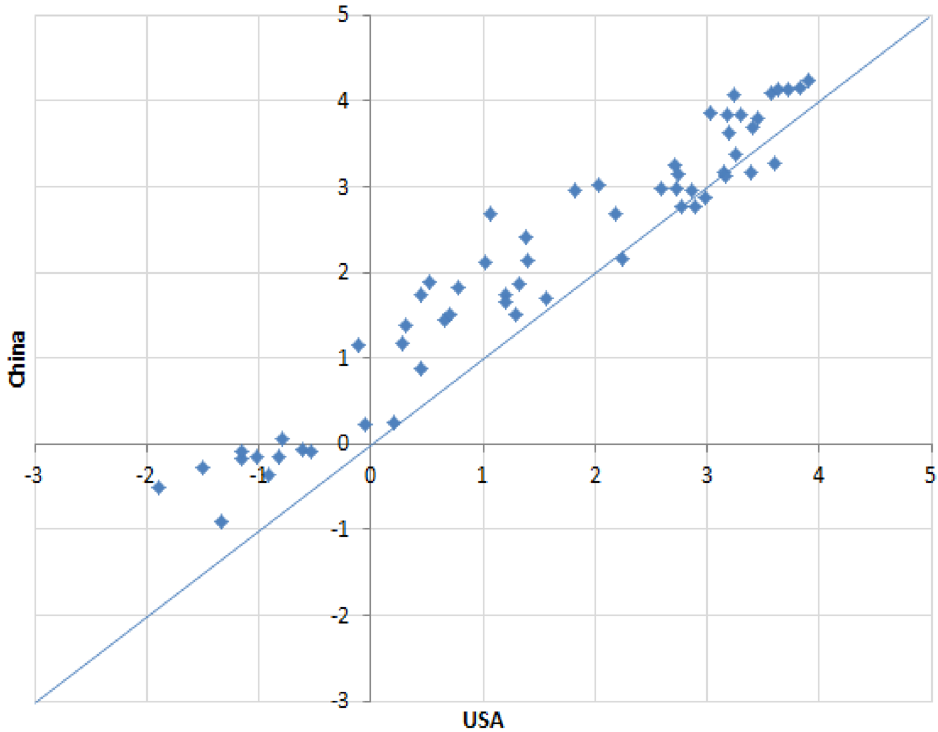


Figure B2-4 Average perceived health in USA and China after information

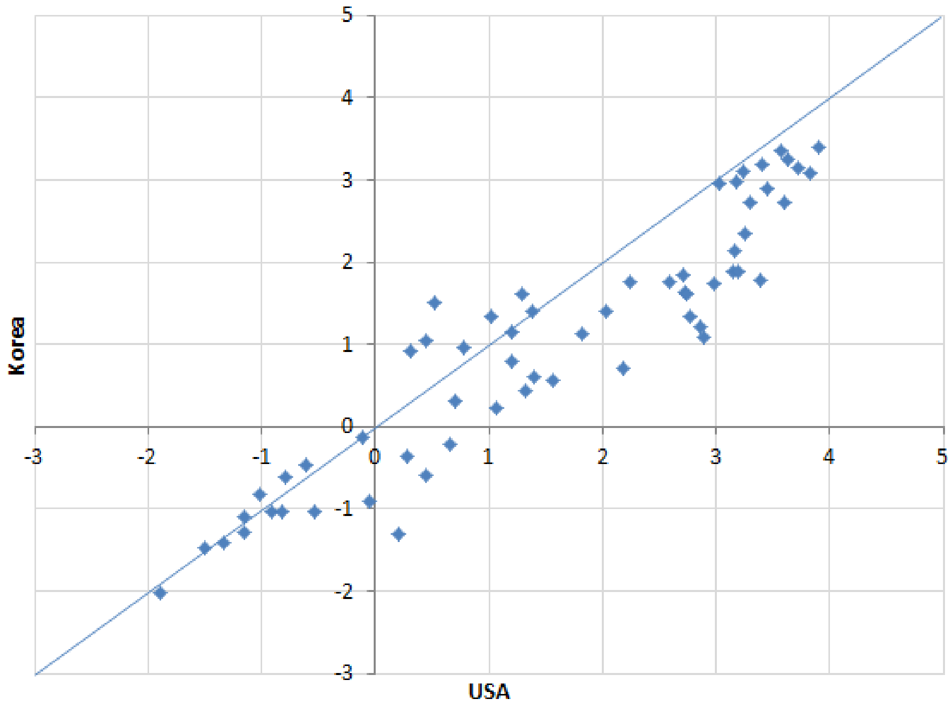


Figure B2-5 Average perceived health in USA and Korea after information

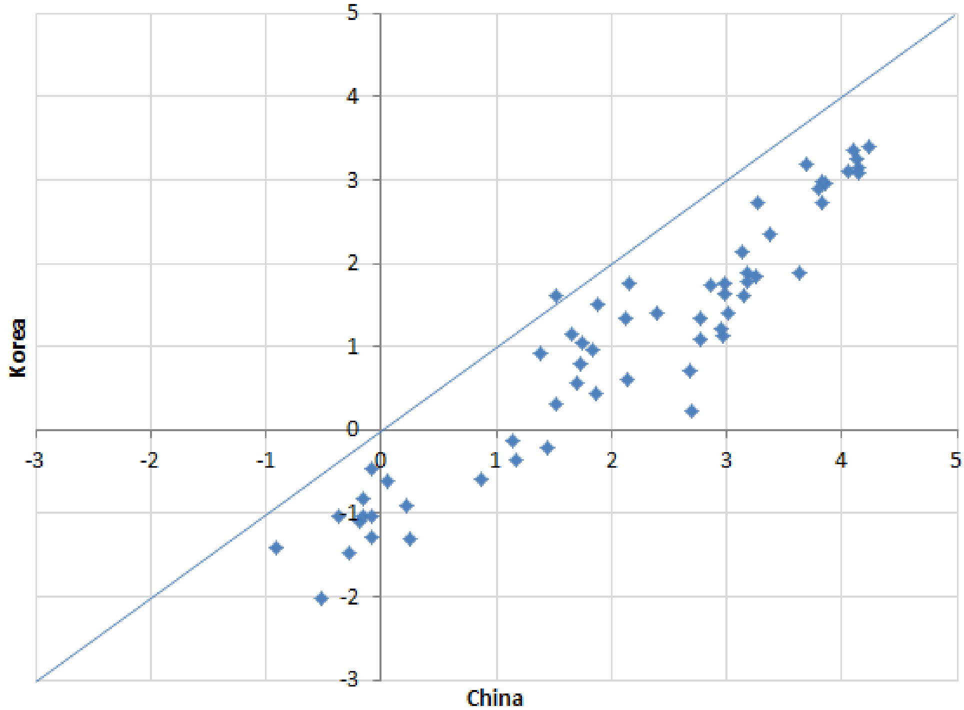


Figure B2-6 Average perceived health in China and Korea after information

Appendix C

Table C2-1 Perceived taste, health, price, and purchase intention for three-cluster model in USA (Before the provision of information)

Cluster	Food Item	Health Signal	Taste	Health	Price	Purchase intention
Ideal	Apple	Green	3.91	4.05	-1.29	3.46
	Banana	Green	3.99	3.95	-1.58	3.73
	Orange	Green	4.02	3.91	-1.19	3.51
	Canned peach	Green	3.16	2.26	-1.24	1.91
	Frozen mixed fruit	Green	3.25	2.89	-0.70	2.49
	Fruit juice	Green	3.79	3.03	-0.96	2.77
	Potato	Green	3.62	2.75	-1.77	3.38
	Lettuce	Green	3.04	3.83	-1.51	3.36
	Tomato	Green	3.26	3.77	-1.35	3.22
	Canned corn	Green	2.88	2.23	-1.54	2.41
	Frozen mixed vegetables	Green	2.49	3.02	-1.41	2.36
	Vegetable juice	Green	2.18	3.69	-0.55	1.74
	Meat-beef	Yellow	3.35	2.01	0.00	2.70
	Meat-chicken	Green	3.43	3.11	-0.99	3.17
	Meat-turkey	Green	3.01	2.99	-0.34	2.18
	Roasted beef	Yellow	2.97	1.69	0.03	1.82
	Salmon	Yellow	2.59	3.28	0.73	1.97
	Tilapia	Green	1.94	2.62	-0.38	1.14
	Catfish	Green	1.83	2.32	-0.04	0.77
	Frozen shrimp	Green	2.68	2.31	0.25	1.79
	Canned tuna	Green	2.10	2.30	-1.38	1.65
	Milk	Green	3.05	3.10	-0.92	3.07
	Cheese	Yellow	3.72	2.27	-0.76	3.24
	Yogurt	Green	2.94	3.12	-1.07	2.23
	Sandwich bread	Yellow	3.15	2.37	-1.20	2.98
	Rice	Yellow	3.09	2.71	-1.62	3.24
	Pasta	Yellow	3.34	1.92	-1.64	2.98
	Cereal	Yellow	3.63	2.07	-0.84	2.93
	Peanut butter	Red	3.40	2.09	-1.27	2.95
	Sandwich	Green	3.73	2.25	-0.85	2.84
	Salad	Green	3.70	3.67	-0.67	3.07
	Soup	Green	3.38	2.89	-1.43	2.85
	Yubari	Green	3.18	3.57	-0.86	2.30
Hedonic	Ground beef	Yellow	3.26	1.38	-0.70	2.74
	Meat-pork	Yellow	2.82	1.37	-0.62	1.71
	Bacon	Red	3.40	0.14	-0.43	2.39
	Sausage	Red	2.90	0.21	-0.41	1.57

	Ham	Green	2.91	1.23	-0.39	1.99
	Ice cream	Yellow	3.95	0.05	-0.32	2.80
	Muffin	Red	3.28	0.68	-1.07	2.07
	Doughnut	Red	3.58	-0.95	-1.43	1.81
	Cookie	Red	3.54	-0.22	-1.17	2.49
	Flour	Yellow	1.90	1.27	-1.29	2.48
	Candy	Red	3.06	-1.51	-1.72	1.36
	Chocolate	Red	3.76	0.92	-1.13	2.70
	Butter	Red	2.86	0.53	-0.96	2.63
	Margarine	Red	1.77	-0.19	-1.20	0.68
	Salad dressing	Red	3.06	0.97	-1.05	2.30
	Hamburger	Yellow	3.71	-0.06	-1.10	2.63
	Pizza	Yellow	3.88	0.47	-0.93	3.28
	Hot dog	Yellow	3.28	-0.69	-1.64	2.07
	Chicken tender	Yellow	3.47	0.75	-1.03	2.50
	French fries	Yellow	3.72	-0.63	-1.47	2.77
	Burrito	Green	3.55	0.69	-1.20	2.41
Uncommon	Frozen scallop	Green	1.42	1.85	0.41	0.22
	Beluga caviar	Yellow	0.19	0.90	2.05	-0.91
	Foie gras	Red	0.60	0.66	0.74	-0.25
	White truffle	Yellow	1.74	1.19	0.96	0.07
	Saffron	Yellow	1.18	1.55	0.56	0.26
	Donkey cheese	Yellow	0.21	0.27	0.19	-0.60

Table C2-2 Perceived taste, health, price, and purchase intention for-five cluster model in USA (After the provision of information)

Cluster	Food Item	Health Signal	Taste	Health	Price	Purchase Intention
Ideal	Apple	Green	3.79	3.91	-1.34	3.44
	Banana	Green	3.85	3.72	-1.52	3.61
	Orange	Green	3.75	3.83	-1.16	3.35
	Canned peach	Green	3.18	2.89	-1.46	2.33
	Frozen mixed fruit	Green	3.29	3.16	-0.72	2.53
	Fruit juice	Green	3.68	3.30	-0.71	2.82
	Potato	Green	3.47	3.03	-1.68	3.23
	Lettuce	Green	3.30	3.64	-1.26	3.46
	Tomato	Green	3.38	3.58	-1.37	3.24
	Canned corn	Green	3.08	2.86	-1.50	2.63
	Frozen mixed vegetables	Green	2.94	3.40	-1.36	2.55
	Vegetable juice	Green	2.67	3.41	-0.57	1.91
	Meat-chicken	Green	3.49	3.25	-0.82	3.09
	Meat-turkey	Green	3.03	3.15	-0.63	2.63
	Canned tuna	Green	2.44	2.98	-1.07	1.99
	Milk	Green	3.19	3.24	-0.98	3.04
	Yogurt	Green	3.00	3.18	-1.16	2.37
	Sandwich	Green	3.56	2.77	-0.82	2.97
	Salad	Green	3.68	3.61	-0.79	3.16
	Soup	Green	3.40	3.19	-1.32	3.04
	Yubari	Green	3.16	3.45	-0.59	2.40
Moderately Ideal	Ground beef	Yellow	3.08	1.20	-0.80	2.55
	Meat-beef	Yellow	3.14	1.38	-0.43	2.40
	Meat-pork	Yellow	2.68	1.02	-0.73	1.82
	Roasted beef	Yellow	2.78	1.20	-0.23	1.80
	Cheese	Yellow	3.31	1.30	-1.17	2.87
	Ice cream	Yellow	3.51	0.29	-0.71	2.63
	Sandwich bread	Yellow	2.86	1.57	-1.35	2.61
	Rice	Yellow	2.99	1.82	-1.55	2.71
	Pasta	Yellow	3.27	1.32	-1.73	2.90
	Flour	Yellow	1.88	1.07	-1.61	2.38
	Cereal	Yellow	3.34	1.39	-1.05	2.64
	Peanut butter	Red	3.00	0.20	-1.54	2.26
	Hamburger	Yellow	3.37	0.45	-1.11	2.59
	Pizza	Yellow	3.58	0.65	-1.20	3.01
	Chicken tender	Yellow	3.22	0.71	-0.99	2.26
	Burrito	Green	3.56	2.03	-1.09	2.71

Taste-oriented	Bacon	Red	2.97	-1.01	-0.78	2.04
	Sausage	Red	2.26	-1.16	-1.03	1.03
	Muffin	Red	2.90	-0.82	-1.24	1.26
	Doughnut	Red	2.97	-1.50	-1.50	1.43
	Cookie	Red	3.31	-1.16	-1.51	2.05
	Candy	Red	2.87	-1.88	-2.00	1.01
	Chocolate	Red	3.27	-0.53	-1.24	2.23
	Butter	Red	2.49	-0.92	-1.40	1.96
	Margarine	Red	1.34	-1.33	-1.48	0.26
	Salad dressing	Red	2.65	-0.61	-1.38	1.68
	Hot dog	Yellow	2.85	-0.12	-1.64	1.84
	French fries	Yellow	3.42	-0.06	-1.52	2.61
Health-oriented	Ham	Green	2.91	2.18	-0.46	2.32
	Salmon	Yellow	2.36	2.25	0.08	1.52
	Tilapia	Green	2.09	2.75	-0.36	1.37
	Catfish	Green	2.03	2.71	-0.37	1.19
	Frozen shrimp	Green	2.49	2.73	0.09	1.85
	Frozen scallop	Green	1.88	2.59	0.56	0.83
Uncommon	Beluga caviar	Yellow	0.23	0.45	1.64	-0.82
	Foie gras	Red	0.14	-0.79	0.29	-0.98
	White truffle	Yellow	1.29	0.53	0.38	0.12
	Saffron	Yellow	1.21	0.78	0.23	-0.09
	Donkey cheese	Yellow	0.55	0.31	0.04	-0.39

Table C2-3 Perceived taste, health, price, and purchase intention for three-cluster model in China (Before the provision of information)

Cluster	Food Item	Health Signal	Taste	Health	Price	Purchase Intention
Ideal	Apple	Green	4.07	4.28	-2.78	4.08
	Banana	Green	4.09	4.15	-2.54	3.83
	Orange	Green	4.15	4.14	-2.62	3.86
	Fruit juice	Green	3.89	3.57	-2.66	3.34
	Potato	Green	3.66	3.68	-3.00	3.68
	Lettuce	Green	3.84	4.20	-2.80	3.89
	Tomato	Green	3.86	4.16	-2.97	3.81
	Vegetable juice	Green	2.95	3.63	-2.41	2.72
	Meat-beef	Yellow	3.71	3.57	-2.04	3.15
	Meat-pork	Yellow	3.45	2.85	-2.49	3.26
	Meat-chicken	Green	3.45	2.98	-2.47	3.09
	Milk	Green	3.79	4.05	-2.73	3.70
	Yogurt	Green	3.99	4.02	-2.55	3.70
	Rice	Yellow	3.89	4.02	-2.87	4.24
	Flour	Yellow	3.68	3.75	-2.66	3.49
	Cereal	Yellow	3.03	3.29	-2.36	2.70
	Salad	Green	3.17	3.05	-2.32	2.46
	Soup	Green	3.73	3.65	-2.56	3.20
	Yubari	Green	3.97	3.88	-2.48	3.36
Health-oriented	Frozen mixed fruit	Green	2.81	2.34	-1.77	1.93
	Canned corn	Green	2.65	1.88	-2.13	1.61
	Frozen mixed vegetables	Green	2.12	1.98	-2.29	1.52
	Ground beef	Yellow	2.43	2.20	-1.69	1.73
	Meat-turkey	Green	2.84	2.53	-1.72	1.96
	Roasted beef	Yellow	3.48	2.45	-1.82	2.32
	Bacon	Red	3.25	2.05	-1.91	2.25
	Ham	Green	3.24	1.59	-2.16	2.30
	Salmon	Yellow	3.36	3.30	-1.45	2.25
	Tilapia	Green	3.01	2.89	-1.59	2.03
	Catfish	Green	2.91	2.99	-1.75	2.17
	Frozen shrimp	Green	2.72	2.04	-1.84	1.97
	Frozen scallop	Green	2.67	1.99	-1.78	1.57
	Canned tuna	Green	2.76	1.94	-1.88	1.66
	Cheese	Yellow	3.20	2.42	-2.04	2.04
	Sandwich bread	Yellow	3.03	2.65	-2.43	2.42
	Muffin	Red	3.26	2.16	-2.38	2.25
	Cookie	Red	3.41	2.04	-2.33	2.53
	Pasta	Yellow	3.07	2.79	-1.87	2.17

	Chocolate	Red	3.55	1.70	-2.38	2.47
	Butter	Red	2.47	1.73	-2.04	1.52
	Salad dressing	Red	2.77	1.94	-2.25	2.10
	Peanut butter	Red	2.99	2.24	-2.17	2.13
	Pizza	Yellow	3.19	1.93	-1.99	2.24
	Sandwich	Green	3.12	1.71	-2.30	2.12
	Burrito	Green	3.19	2.17	-2.00	2.30
	Beluga caviar	Yellow	2.72	2.75	-0.48	0.92
	Foie gras	Red	2.89	2.34	-1.28	1.59
	White truffle	Yellow	2.81	2.96	-0.74	1.13
	Saffron	Yellow	1.84	2.49	-1.34	1.44
	Donkey cheese	Yellow	1.86	1.90	-1.43	0.89
Taste-oriented	Canned peach	Green	2.63	1.35	-2.28	1.30
	Sausage	Red	3.16	1.09	-2.33	2.17
	Ice cream	Yellow	3.75	0.80	-2.37	2.76
	Doughnut	Red	2.84	0.94	-2.38	1.73
	Candy	Red	2.59	0.41	-2.66	1.47
	Margarine	Red	1.86	0.28	-2.07	0.48
	Hamburger	Yellow	3.05	0.59	-2.32	2.10
	Hot dog	Yellow	3.10	0.95	-2.32	1.80
	Chicken tender	Yellow	3.20	1.45	-2.37	2.24
	French fries	Yellow	3.16	-0.30	-2.47	1.73

Table C2-4 Perceived taste, health, price, and purchase intention for six-cluster model in China (After the provision of information)

Cluster	Food Item	Health Signal	Taste	Health	Price	Purchase Intention
Ideal	Apple	Green	4.12	4.24	-2.90	4.12
	Banana	Green	4.01	4.14	-2.92	3.92
	Orange	Green	4.10	4.15	-2.91	3.96
	Fruit juice	Green	3.88	3.83	-2.74	3.62
	Potato	Green	3.80	3.86	-2.98	3.89
	Lettuce	Green	3.89	4.13	-2.90	3.94
	Tomato	Green	3.89	4.10	-2.82	3.82
	Vegetable juice	Green	3.46	3.70	-2.81	3.26
	Milk	Green	3.93	4.06	-2.81	3.80
	Yogurt	Green	3.91	3.83	-2.75	3.75
	Soup	Green	3.72	3.63	-2.75	3.28
	Yubari	Green	3.90	3.80	-2.54	3.64
Moderately Ideal	Canned peach	Green	3.12	2.77	-2.49	2.37
	Frozen mixed fruit	Green	3.12	3.13	-2.47	2.50
	Canned corn	Green	3.01	2.95	-2.47	2.39
	Frozen mixed vegetables	Green	2.85	3.17	-2.51	2.52
	Meat-beef	Yellow	3.40	2.40	-2.32	2.70
	Meat-chicken	Green	3.53	3.38	-2.78	3.38
	Meat-turkey	Green	3.17	3.18	-2.15	2.53
	Ham	Green	3.43	2.69	-2.30	2.82
	Tilapia	Green	3.16	3.14	-1.92	2.45
	Catfish	Green	3.09	3.25	-2.38	2.58
	Frozen shrimp	Green	3.03	2.98	-2.29	2.48
	Frozen scallop	Green	3.09	2.98	-2.17	2.35
	Canned tuna	Green	3.14	2.86	-2.08	2.21
	Rice	Yellow	3.62	2.96	-2.81	3.72
	Flour	Yellow	3.46	2.69	-2.76	3.16
	Sandwich	Green	3.42	2.78	-2.48	2.80
	Salad	Green	3.44	3.27	-2.51	2.93
	Burrito	Green	3.37	3.02	-2.40	2.68
Taste-oriented	Ground beef	Yellow	2.71	1.66	-2.09	1.74
	Meat-pork	Yellow	3.30	2.13	-2.69	2.96
	Roasted beef	Yellow	3.19	1.74	-2.21	2.17
	Cheese	Yellow	2.66	1.52	-2.15	1.64
	Ice cream	Yellow	3.29	1.17	-2.56	2.43
	Sandwich bread	Yellow	2.84	1.70	-2.45	2.04
	Pasta	Yellow	2.92	1.87	-2.26	2.05

	Cereal	Yellow	2.94	2.14	-2.42	2.53
	Hamburger	Yellow	2.78	0.87	-2.51	1.84
	Pizza	Yellow	3.08	1.45	-2.10	2.12
	Hot dog	Yellow	2.91	1.15	-2.39	1.98
	Chicken tender	Yellow	3.03	1.52	-2.44	2.08
Less Taste-oriented	Bacon	Red	2.61	-0.16	-2.35	1.43
	Sausage	Red	2.85	-0.18	-2.46	1.71
	Muffin	Red	2.68	-0.16	-2.50	1.26
	Doughnut	Red	2.73	-0.27	-2.49	1.20
	Cookie	Red	2.64	-0.09	-2.52	1.39
	Chocolate	Red	2.95	-0.09	-2.42	1.87
	Salad dressing	Red	2.44	-0.08	-2.31	1.22
	Peanut butter	Red	2.36	0.25	-2.39	1.18
	French fries	Yellow	2.90	0.23	-2.75	1.71
	Foie gras	Red	2.56	0.05	-1.26	0.38
Uncommon	Salmon	Yellow	3.18	2.16	-1.71	1.85
	Beluga caviar	Yellow	2.72	1.75	-0.93	1.05
	White truffle	Yellow	2.78	1.88	-1.02	1.02
	Saffron	Yellow	2.17	1.83	-1.87	1.40
	Donkey cheese	Yellow	1.98	1.38	-1.62	0.82
Unfavorable	Candy	Red	2.45	-0.52	-2.69	0.90
	Butter	Red	2.15	-0.36	-2.22	0.88
	Margarine	Red	1.60	-0.91	-2.18	-0.01

Table C2-5 Perceived taste, health, price, and purchase intention for six-cluster model in Korea (Before the provision of information)

Cluster	Food Item	Health Signal	Taste	Health	Price	Purchase Intention
Ideal	Apple	Green	3.22	3.42	1.30	2.85
	Banana	Green	3.14	2.99	0.52	2.61
	Orange	Green	3.12	3.04	1.14	2.61
	Fruit juice	Green	3.16	2.23	1.56	2.18
	Potato	Green	2.75	2.83	-0.26	2.57
	Lettuce	Green	2.37	3.06	0.19	2.57
	Tomato	Green	2.55	3.58	0.59	2.70
	Meat-chicken	Green	2.83	2.13	0.68	2.21
	Milk	Green	2.60	2.94	1.23	2.44
	Cheese	Yellow	2.70	2.31	1.52	2.00
	Yogurt	Green	3.07	2.85	1.35	2.38
	Salad	Green	2.79	2.76	1.51	1.97
Taste-oriented	Frozen mixed fruit	Green	1.97	1.41	1.20	0.86
	Ground beef	Yellow	2.41	1.40	2.13	1.42
	Meat-pork	Yellow	3.11	1.58	1.26	2.42
	Roasted beef	Yellow	2.70	0.80	2.14	1.28
	Frozen shrimp	Green	1.94	0.99	1.21	1.01
	Canned tuna	Green	2.63	1.42	1.12	1.84
	Sandwich bread	Yellow	2.55	0.74	0.26	2.07
	Rice	Yellow	2.56	1.23	0.63	2.44
	Pasta	Yellow	2.21	0.60	0.57	1.27
	Cereal	Yellow	2.45	0.80	1.02	1.32
	Salad dressing	Red	2.49	0.97	1.39	1.35
	Sandwich	Green	3.23	0.79	1.48	1.88
	Burrito	Green	2.60	0.74	1.31	1.31
	Soup	Green	2.44	1.47	0.78	1.29
Hedonic	Canned peach	Green	2.48	-0.08	0.35	0.70
	Bacon	Red	2.78	-0.06	1.64	1.24
	Sausage	Red	2.94	-0.40	1.27	1.56
	Ham	Green	2.91	-0.40	1.59	1.55
	Ice cream	Yellow	3.32	-0.82	1.53	1.90
	Muffin	Red	2.51	-0.19	1.15	1.08
	Doughnut	Red	2.71	-1.07	0.72	0.85
	Cookie	Red	2.91	-0.46	0.94	1.26
	Chocolate	Red	2.97	0.09	0.86	1.44
	Butter	Red	2.23	-0.06	1.20	0.92
	Peanut butter	Red	1.98	-0.41	0.98	0.38
	Hamburger	Yellow	3.11	-1.16	0.92	1.64
	Pizza	Yellow	3.10	-0.83	2.00	1.76

	Hot dog	Yellow	2.91	-0.57	0.61	1.38
	Chicken tender	Yellow	2.96	-0.06	1.11	1.64
Health-oriented	Vegetable juice	Green	1.58	3.17	1.59	1.74
	Meat-beef	Yellow	3.03	1.85	2.83	1.83
	Salmon	Yellow	2.41	2.56	2.44	1.49
	Beluga caviar	Yellow	1.28	1.67	3.18	-0.64
	White truffle	Yellow	1.32	2.68	3.24	-0.38
	Yubari	Green	2.77	2.57	2.40	1.64
Less Taste-oriented	Canned corn	Green	2.20	0.20	-0.19	0.89
	Flour	Yellow	2.11	-0.20	-0.23	1.65
	Candy	Red	2.09	-1.85	-1.10	-0.30
	Margarine	Red	1.42	-0.89	-0.13	-0.04
	French fries	Yellow	2.71	-1.52	-0.26	1.07
Less Health-oriented	Frozen mixed vegetables	Green	0.72	0.91	0.70	0.01
	Meat-turkey	Green	1.34	1.44	1.87	0.23
	Tilapia	Green	0.48	0.95	1.21	-0.40
	Catfish	Green	1.07	1.70	1.57	0.26
	Frozen scallop	Green	1.29	1.06	1.29	0.33
	Foie gras	Red	0.34	0.78	2.44	-1.03
	Saffron	Yellow	0.34	1.27	1.79	-0.65
	Donkey cheese	Yellow	0.43	1.22	1.62	-0.31

Table C2-6 Perceived taste, health, price, and purchase intention for three-cluster model in Korea (After the provision of information)

Cluster	Food Item	Health Signal	Taste	Health	Price	Purchase Intention
Ideal	Apple	Green	3.25	3.39	1.16	2.89
	Banana	Green	3.29	3.15	0.54	2.80
	Orange	Green	3.23	3.07	1.18	2.69
	Fruit juice	Green	3.12	2.73	1.59	2.17
	Potato	Green	2.87	2.96	0.03	2.66
	Lettuce	Green	2.68	3.25	0.08	2.64
	Tomato	Green	2.68	3.36	0.61	2.58
	Vegetable juice	Green	2.06	3.19	1.59	1.73
	Meat-chicken	Green	2.94	2.34	0.78	2.48
	Milk	Green	2.71	3.10	1.22	2.61
	Yogurt	Green	3.11	2.98	1.34	2.53
	Salad	Green	2.88	2.72	1.40	2.17
	Yubari	Green	2.85	2.90	2.32	1.72
Health-oriented	Canned peach	Green	2.47	1.09	0.53	1.19
	Frozen mixed fruit	Green	2.07	2.14	1.48	1.46
	Canned corn	Green	2.43	1.21	0.08	1.31
	Frozen mixed vegetables	Green	1.38	1.77	0.96	0.78
	Ground beef	Yellow	2.35	1.15	1.95	1.42
	Meat-beef	Yellow	2.68	1.41	2.58	1.64
	Meat-pork	Yellow	2.83	1.34	1.20	2.21
	Meat-turkey	Green	1.89	1.89	2.04	0.59
	Roasted beef	Yellow	2.56	0.79	2.14	1.15
	Ham	Green	2.78	0.70	1.48	1.67
	Salmon	Yellow	2.38	1.76	2.28	1.18
	Tilapia	Green	0.96	1.60	1.41	0.02
	Catfish	Green	1.31	1.84	1.59	0.38
	Frozen shrimp	Green	2.20	1.64	1.54	1.40
	Frozen scallop	Green	1.72	1.76	1.57	0.73
	Canned tuna	Green	2.55	1.73	0.98	1.85
	Cheese	Yellow	2.49	1.61	1.46	1.79
	Sandwich bread	Yellow	2.48	0.57	0.35	1.77
	Rice	Yellow	2.36	1.12	0.52	2.24
	Pasta	Yellow	2.00	0.43	0.62	1.07
	Cereal	Yellow	2.32	0.61	0.95	1.22
	Chicken tender	Yellow	2.71	0.31	1.21	1.50
	Sandwich	Green	2.80	1.34	1.45	2.13
	Burrito	Green	2.64	1.40	1.52	1.64
	Soup	Green	2.57	1.89	0.98	1.69

	Beluga caviar	Yellow	1.13	1.05	2.81	-0.71
	White truffle	Yellow	1.46	1.52	2.73	-0.32
	Saffron	Yellow	0.78	0.96	1.84	-0.49
	Donkey cheese	Yellow	1.04	0.92	1.73	-0.05
Taste-oriented	Bacon	Red	2.38	-0.82	1.42	0.59
	Sausage	Red	2.65	-1.09	1.19	1.07
	Ice cream	Yellow	2.88	-0.36	1.37	1.53
	Muffin	Red	2.28	-1.04	0.85	0.41
	Doughnut	Red	2.54	-1.48	0.69	0.63
	Cookie	Red	2.54	-1.28	0.62	0.58
	Flour	Yellow	2.14	0.23	-0.07	1.47
	Candy	Red	1.90	-2.03	-1.17	-0.58
	Chocolate	Red	2.63	-1.04	0.83	1.07
	Butter	Red	1.87	-1.03	0.98	0.28
	Margarine	Red	1.17	-1.41	0.02	-0.29
	Salad dressing	Red	2.17	-0.46	1.02	0.57
	Peanut butter	Red	1.72	-1.30	0.65	-0.18
	Hamburger	Yellow	2.74	-0.60	0.93	1.46
	Pizza	Yellow	2.79	-0.21	1.69	1.52
	Hot dog	Yellow	2.77	-0.13	0.83	1.35
	French fries	Yellow	2.55	-0.92	0.07	1.05
	Foie gras	Red	0.29	-0.60	2.29	-1.40

Appendix D

Oklahoma State University Institutional Review Board

Date:	Thursday, July 14, 2016
IRB Application No	AG1620
Proposal Title:	Consumers' perception on taste, health, and expense of different foods
Reviewed and Processed as:	Exempt

Status Recommended by Reviewer(s): Approved Protocol Expires: 7/13/2019

Principal Investigator(s):

Jisung Jo	Jayson Lusk
	411 Ag Hall
Stillwater, OK 74078	Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

- The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval. Protocol modifications requiring approval may include changes to the title, PI advisor, funding status or sponsor, subject population composition or size, recruitment, inclusion/exclusion criteria, research site, research procedures and consent/assent process or forms
2. Submit a request for continuation if the study extends beyond the approval period. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of the research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Dawnett Watkins 219 Scott Hall (phone: 405-744-5700, dawnett.watkins@okstate.edu).

Sincerely,

Hugh Crethar, Chair
Institutional Review Board

Thank you for participating in this study. The following contains information about your study and your rights as a research participant.

Project Title: Consumers' perception on taste, health, and expense of different foods

Investigator: Jisung Jo, Oklahoma State University

Purpose: This is a web-based survey research study designed to investigate consumers' taste, health, and price perceptions for 60 food items.

Procedures: Proceeding with the web-based survey will imply your consent to participate in this study. There are about 143 questions asking about your perception for different food items in addition to questions asking about food values. We also ask some basic demographic questions. The survey will take about 30 minutes to complete.

Risks of Participation: The risks associated with this study are minimal. The risks are not greater than those ordinarily encountered in daily life. Moreover, you may stop the survey at any time.

Benefits: This research will assist researchers understand why people buy different food and how food choices are affected by taste, health, and price.

Confidentiality: The researchers will not have access to your name. At no point will a data file be constructed in which your name is linked with your responses. The data will be stored by the principal investigators in their office with no intention to destroy the data. The data will only be released in summaries in which no individual's answers can be identified.

Contacts: If you have any questions or concerns about this project, please contact Jisung Jo, jisung.jol@okstate.edu, 405-385-3184 or Jayson L. Lusk, jayson.lusk@okstate.edu or 405-744-7465. If you have questions about your rights as a research volunteer, you may contact Dr. Hugh Crethar, IRB Chair at 223 Scott Hall, Stillwater, OK 74078, 405-744-3377 or irb@okstate.edu.

Participant Rights: Your participation in this research in voluntary. You can discontinue the survey at any time without reprisal or penalty.

Consent: I have read and fully understand the consent form. I understand that my participation is voluntary. By clicking below, I am indicating that I freely and voluntarily and agree to participate in this study and I also acknowledge that I am at least 18 years of age.

It is recommended that you print a copy of this consent page for your records before you begin.

CHAPTER III

PREDICTING FOOD PRICES USING DATA FROM CONSUMER SURVEY AND SEARCH

Introduction

Although food comprises a relatively small share of consumers' budgets, changes in food prices can have an important impact on household well-being, particularly for lower-income consumers who spend a larger portion of their income on food than higher-income consumers. In fact, many economic analysts focus only on the "core" consumer price index (CPI), which excludes food and energy prices, because of a belief that prices for food and energy are "volatile and are subject to price shocks that cannot be damped through monetary policy" (Greenlees and McClelland, 2008). Coupling food price volatility with the fact that food is purchased frequently implies that consumers may be more aware of or attentive to changes in the price of food than with other items. In fact, the data suggest low-income households tend to pay less for the same food items than the rich, perhaps because of greater price sensitivity and search behavior (Broda et al., 2009). As such, data related to consumers' price knowledge and expectations may be useful in forecasting changes in the price of food.

Projecting food prices is of interest to participants of the food supply chain as well as government agencies. Firms make production decisions based on price expectations, and agribusiness firms hedge commodity and output prices based on expected prices. Moreover, changing food prices have implications for a number of government programs such as the supplemental nutritional assistance program (SNAP), the women, infants, and children (WIC) program, and the school lunch program, among others. Because of the desire to anticipate future food prices, a number of ongoing efforts exist to forecast the food component of the CPI (e.g. Kuhns et al., 2015).

Virtually all existing efforts to forecast the food-related CPI rely on time series models where future price changes are estimated as a function of past food prices and lagged values of related variables (Joutz, 1997). These models are thus backward looking. However, a number of more forward-looking variables are available that might be useful in predicting food price changes. In this paper, we consider two such measures: a survey-based index (the Index of Consumer Sentiment (ICS) from the University of Michigan) and a search-based Google Trends Index (GTI).

Previous research suggests the potential for survey-based sentiment indices like the ICS to forecast future food prices, even though ICS reflects overall sentiment not just focused on food. Wilcox (2007) found that inclusion of the ICS in a model improved forecasts of consumption and expenditures on durable as well as non-durable goods and services. Ang, Bekaert, and Wei (2007) also found that survey forecasts outperform other forecasts based on time series models, an economic model of the Philips curve, and information embedded in asset prices. Girardi, Gayer, and Reuter (2015) also found survey data to be useful in forecasting economic growth measures. They highlight the utility of using survey data for "nowcasting" given that releases of public data, such as the CPI, often occur with a significant lag.

In addition to survey-based measures, newer measures related to consumers' Internet search behavior are now available. According to the World Bank data, internet users in 2014 represent 87.36% of the United States of America's population. Prior research has shown some promise in using measures like the Google Trends search-based index as a leading indicator of private consumption (Choi and Varian, 2012; Ginsberg et al, 2009; Souchoy, 2009; and Vosen and Schmidt, 2011). Swallow and Labbe (2013) show that Google Trends search results provide the most useful information about sales of automobiles in an emerging market. They show that the models incorporating the Google Trends Automotive Index outperform benchmark specifications for both in-sample and out-of-sample nowcasts. Further, Vosen and Schmidt (2011) compared the Google Trends search-based index to a survey-based indicies, such as the Index of Consumer Sentiment from Michigan survey and the Consumer Confidence Index from the Conference Board, and found that all of the Google Trends indicators outperform the survey-based indicators in terms of forecast performance.

In this research, we explore whether ICS and GTI improve the performance of Food and Beverage CPI forecast models. Moreover, we compare the forecast performance of our models utilizing ICS and GTI data with the forecasts released by the USDA Economic Research Service. We find that not only are consumers' price expectation indices meaningful determinants of future food price changes but that models incorporating these measures outperform USDA forecasts.

Data

Food-Related Consumer Price Index

The U.S. Bureau of Labor Statistics (BLS) reports the Consumer Price Index (CPI) as an economic indicator, a deflator of other economic series, and a means of adjusting dollar values. The CPI represents the average change in prices paid by urban consumers for a market basket of goods and services over time. Urban consumers are divided into two groups: all urban consumers and urban wage earners and clerical workers. The first group covers 87 percent of the total U.S.

population and includes professionals, the self-employed, the poor, and the unemployed. Because the subjects of this group are residents of a metropolitan area, the Consumer Price Index for all urban consumers (CPI-U) does not reflect the spending patterns of people who live in rural nonmetropolitan areas, such as farm families. The Consumer Price Index for urban wage earners and clerical workers (CPI-W) is the index based on the second group. To be considered as a member of the second group, more than one-half of the household's income must come from clerical or wage occupations and at least one of the household's earners must have been employed for at least 37 weeks of the last 12 months. As a subset of the first group, the second group covers around 32 percent of the U.S population.

The market basket of goods and services reflected in the CPI can be separated into eight categories: food and beverages, housing, apparel, transportation, medical care, recreation, education and communication, and other goods and services. From 2011 to 2012, the relative importance of the food and beverage component in the CPI-U was 14.9 out of 100. This research investigates the movement of the Food and Beverages CPI-U with reference base, 1982-84=100. We also investigate whether the total CPI across eight categories is an exogenous predictor of the Food and Beverages CPI.

Figure 3-1 shows that both the total CPI and Food and Beverages CPI trended upward from 2004 to 2015. During the periods between 2008 and 2009, while the Food and Beverages CPI and the total CPI moved in opposite directions, it is perhaps as a result of monetary policy associated with the Great Recession. These price movements support Greenlees and McClelland's (2008) argument that food price shocks cannot be damped through monetary policy. Including data from the financial crisis period in the forecasting model is thus necessary to understand more about the structural relationship and long-run dynamic behavior of multivariate time series. We hypothesize that a vector error correction model (VECM) will outperform other forecasting models because the error correction term could capture how the variables react when they move out of long-run equilibrium (Zivot and Wang, 2007).

Consumer Sentiment

Several survey-based indices of consumer sentiment are available, such as the Livingston survey and the Survey of Professional Forecasters (SPF). These indices are provided twice a year, in June and December, and the middle of every quarter, respectively. Both of these measures are based on surveys of economists from industry, government, and academia. Unlike the Livingston and SPF, the Index of Consumer Sentiment from Michigan is measured monthly and participants are households. As such, the ICS is likely to be a more appropriate index to apply consumers' expectations and sentiment to forecast food-related CPI.

The University of Michigan has reported monthly ICS data since 1978, and the reference base is March 1997. The ICS is derived from the following five questions:

Q_1. Personal Finance Current: We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?

Q_2. Personal Finance Expected: Now looking ahead—do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?

Q_3. Business Condition 12 Month: Now turning to business conditions in the country as a whole—do you think that during the next twelve months we will have good times financially, or bad times, or what?

Q_4. Business Condition 5 years: Looking ahead, which would you say is more likely—that in the country as a whole we will have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?

Q_5. Buying Conditions: About the big things people buy for their homes, such as furniture, a refrigerator, stove, television, and things like that—generally speaking, do you think now is a good or bad time for people to buy major household items?

Figure 3-1 shows that the ICS has a cyclical pattern. Between 2007 and 2008, which is the beginning of the financial crisis in the U.S., consumer sentiment fell and has, in more recent months begun to rise.

Search-Based Index (Google Trends Index)

Google Trends provides a measure of the popularity of terms for which Google users have searched over time. The index of Google Trends measures the number of searches conducted for a particular term, relative to the total number of searches done on Google over time. Specifically,

(1) $$Google\ trends\ A_t = \frac{SA_t}{\max(\mathrm{SA_1,SA_2,\ldots,SA_t})} \times 100,$$

where $Google\ trends\ A_t$ is a percentage of a certain term entered at *t*-th period, SA_t is the absolute search numbers of term A at *t*-th period, and $\max(\mathrm{SA_1, SA_2, \ldots, SA_t})$ is the highest values among SA_t. $Google\ trends\ A_t$ is presented on a scale from 0 to 100. In this study, we create an index based on the search term “food prices.” The Google Trends Index is available from January 2004, and the highest point in our data is May 2008.

In the long run, the GTI has a cyclical (or nonlinear) pattern like the ICS. As can be seen from Figure 3-1, the ICS and GTI have different structures, especially during the financial crisis, which also coincided with a time of high agricultural commodity prices. Results suggest people searched more frequently for, or are more worried about, the price of necessities during the period of economic instability.

Methods

To construct the consumer-oriented Food and Beverages CPI forecast model, we perform several tests. First, the ADF unit root test is conducted to investigate the variables' stationarity over time. This is also the first step of the cointegration rank test. Second, to determine the exogenous variables for a vector autoregression with exogenous variables (VAR-X) and a vector error correction model with exogenous variables (VECM-X), the weak exogeneity test and the Granger causality test are applied. Third, by conducting the cointegration rank test between variables, we obtain the long-run equilibrium structure between endogenous variables. Also, this test will be used for the vector error correction model (VECM) and a VECM-X model. Fourth, we evaluate alternative forecasting models with both a moving window and an expanding window scheme. Lastly, to compare the conventional forecast from the USDA with the consumer-oriented forecast model, an encompassing test is used.

ARIMAX model

While the pure autoregressive integrated moving average (ARIMA) model is composed of lagged dependent variables and errors, an autoregressive integrated moving average model with exogenous variables (ARIMA-X) includes the dependent variable, lagged dependent variable, and the other variables in the equation to explain the external effect on the dependent variables. The ARIMA-X model assumes that the future value of a variable is a linear function of past observations and independent variables. The general ARIMA-X (*p,d,q*) process has the form:

$$(2) \qquad \Delta y_t = \theta_0 + \sum_{i=1}^{p} \emptyset_i \Delta y_{t-i} + \varepsilon_t - \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \sum_{j=1}^{s} \pi_j \Delta x_{jt-1},$$

where Δy_t is the differenced time series values at time *t*, Δy_{t-i} denotes the differenced previous values at time *t-i*, ε_t is random error which follows a white noise process, Δx_{jt-1} is the *j th* independent variable at time *t*-1, *p* is the number of autoregressive terms, *q* is the number of moving-average terms, and s is the number of exogenous variables.

In this research, the CPI of all items (AllCPI), the Google Trend Index (GTI), and the Index of Consumer Sentiment (ICS) are considered as exogenous variables. Thus, the first specifications of the ARIMA-X (*p,d,q*) model are:

$$(3) \qquad \Delta lnFCPI_t = \theta_0 + \sum_{i=1}^{p} \emptyset_i \, \Delta lnFCPI_{t-i} + \theta_1 \Delta lnAllCPI_{t-1} + \theta_2 \Delta lnGTI_{t-1} +$$

$$\theta_3 \Delta lnICS_{t-1} + +\varepsilon_t - \sum_{k=1}^{q} \rho_j \varepsilon_{t-j},$$

where $\Delta lnFCPI_t$ is the first differenced Food and Beverages category's Consumer Price Index, $\Delta lnFCPI_{t-i}$ is the first differenced *i th* lags of $\Delta lnFCPI_t$, $\Delta lnAllCPI_{t-1}$ is the first differenced Consumer Price Index about all items at time *t*-1, $\Delta lnGTI_{t-1}$ is the first differenced Google Trends Index about "Food Prices" at time *t*-1, $\Delta lnICS_{t-1}$ is the first differenced Index of Consumer Sentiment at time *t*-1, and ε_t is the stochastic error term which is independently and identically distributed with a mean of zero and constant variance of σ^2.

VAR and VARX models

A vector autoregression (VAR) model is a multivariate extension of the simple autoregressive model. Sims (1980) proposed models where all variables are jointly endogenous. The main goal of the VAR model is to determine the interrelationship among variables. Thus, Sims (1980) and Sims, Stock, and Watson (1990) suggest the variables in levels are more appropriate than those of differencing, even if the variables are not stationary over time. Of course, the VAR in first differences is possible. The VAR(p) model in standard form is:

$$(4) \qquad x_t = A_0 + \sum_{i=1}^{p} A_i x_{t-i} + e_t,$$

where x_t is a (n×1) vector containing each of the n variables included in the VAR, A_0 is a (n×1) vector of intercept terms, A_i is (n×n) matrices of coefficients, and e_t is a (n×1) vector of error terms.

Now consider a VAR(p) in levels:

$$(5) \qquad \begin{bmatrix} lnFCPI_t \\ lnAllCPI_t \\ lnGTI_t \\ lnICS_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} \alpha_{11}^1 & \alpha_{12}^1 & \alpha_{13}^1 & \alpha_{14}^1 \\ \alpha_{21}^1 & \alpha_{22}^1 & \alpha_{23}^1 & \alpha_{24}^1 \\ \alpha_{31}^1 & \alpha_{32}^1 & \alpha_{33}^1 & \alpha_{34}^1 \\ \alpha_{41}^1 & \alpha_{42}^1 & \alpha_{43}^1 & \alpha_{44}^1 \end{bmatrix} \begin{bmatrix} lnFCPI_{t-1} \\ lnAllCPI_{t-1} \\ lnGTI_{t-1} \\ lnICS_{t-1} \end{bmatrix} + \cdots +$$

$$\begin{bmatrix} \alpha_{11}^p & \alpha_{12}^p & \alpha_{13}^p & \alpha_{14}^p \\ \alpha_{21}^p & \alpha_{22}^p & \alpha_{23}^p & \alpha_{24}^p \\ \alpha_{31}^p & \alpha_{32}^p & \alpha_{33}^p & \alpha_{34}^p \\ \alpha_{41}^p & \alpha_{42}^p & \alpha_{43}^p & \alpha_{44}^p \end{bmatrix} \begin{bmatrix} lnFCPI_{t-p} \\ lnAllCPI_{t-p} \\ lnGTI_{t-p} \\ lnICS_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{FCPIt} \\ \varepsilon_{AllCPIt} \\ \varepsilon_{GTIt} \\ \varepsilon_{ICSt} \end{bmatrix},$$

where α_{ij}^k i = 1,2,3,4, j=1,2,3,4 and k=1,2,…p, are the autoregressive coefficients and ε_{FCPIt}, $\varepsilon_{AllCPIt}$, ε_{GTIt}, and ε_{ICSt} are white-noise disturbances with standard deviations of σ_{FCPI}, σ_{AllCPI}, σ_{GTI}, and σ_{ICS}, respectively.

To determine the exogenous variables for the vector autoregressive model with the exogenous variable (VAR-X), the weak exogeneity test and Granger-causality test are conducted. The standard VAR-X model is

$$(6) \qquad x_t = A_0 + \sum_{i=1}^{p} A_i x_{t-i} + \sum_{i=1}^{q} B_i y_{t-i} + e_t,$$

where y_t is a (n×1) vector of exogenous variables, B_i is (n×n) matrices of coefficients, and e_t is a vector of error terms.

VECM and VECMX models

A vector error-correction (VECM) model indicates how short-term dynamics of variables in the system are influenced by discrepancies from long-run equilibrium. In the equation, each variable in the left hand side responds to the previous period's deviation from long-run equilibrium, their own and others' lagged values, and white noise process. Because the left side of the equation is I(0), the right hand side should be I(0). That is, the linear combination of endogenous variables must be stationary. The generalized n-variable VECM model is:

$$(7) \qquad \Delta x_t = A + \Pi x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t,$$

where A is a (n×1) vector of intercept terms with elements A_j, j=1,2,3,…,n; ϕ_i is a (n×n) coefficient matrices with elements $\phi_{jk}(i)$, k=1,2,3,…n; Π is a matrix with elements $\alpha\beta'$, where α is the speed of adjustment coefficients and β is the long-run parameters; and e_t is a (n×1) vector with elements e_{it}.

As specified, the VECM model form for this research is:

$$(8) \qquad \begin{bmatrix} \Delta lnFCPI_t \\ \Delta lnAllCPI_t \\ \Delta lnGTI_t \\ \Delta lnICS_t \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{bmatrix} \begin{bmatrix} lnFCPI_{t-1} \\ lnAllCPI_{t-1} \\ lnGTI_{t-1} \\ lnICS_{t-1} \end{bmatrix} +$$

$$\begin{bmatrix} \varphi_{11}^1 & \varphi_{12}^1 & \varphi_{13}^1 & \varphi_{14}^1 \\ \varphi_{21}^1 & \varphi_{22}^1 & \varphi_{23}^1 & \varphi_{24}^1 \\ \varphi_{31}^1 & \varphi_{32}^1 & \varphi_{33}^1 & \varphi_{34}^1 \\ \varphi_{41}^1 & \varphi_{42}^1 & \varphi_{43}^1 & \varphi_{44}^1 \end{bmatrix} \begin{bmatrix} \Delta lnFCPI_{t-1} \\ \Delta lnAllCPI_{t-1} \\ \Delta lnGTI_{t-1} \\ \Delta lnICS_{t-1} \end{bmatrix} + \cdots +$$

$$\begin{bmatrix} \varphi_{11}^{p-1} & \varphi_{12}^{p-1} & \varphi_{13}^{p-1} & \varphi_{14}^{p-1} \\ \varphi_{21}^{p-1} & \varphi_{22}^{p-1} & \varphi_{23}^{p-1} & \varphi_{24}^{p-1} \\ \varphi_{31}^{p-1} & \varphi_{32}^{p-1} & \varphi_{33}^{p-1} & \varphi_{34}^{p-1} \\ \varphi_{41}^{p-1} & \varphi_{42}^{p-1} & \varphi_{43}^{p-1} & \varphi_{44}^{p-1} \end{bmatrix} \begin{bmatrix} \Delta lnFCPI_{t-p-1} \\ \Delta lnAllCPI_{t-p-1} \\ \Delta lnGTI_{t-p-1} \\ \Delta lnICS_{t-p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{FCPIt} \\ \varepsilon_{AllCPIt} \\ \varepsilon_{GTIt} \\ \varepsilon_{ICSt} \end{bmatrix}$$

The VECM model can be expressed with a multivariate VAR model in first differences augmented by the error correction term when $\gamma_{ij} = 0$. Therefore, at least one γ_{ij} should not be zero. Like the VAR-X model, the weak-exogenous test and Granger-causality test are used to determine exogenous variables for a vector error correction model with exogenous variable (VECM-X) model. The generalized form of the VECM-X model is:

(9) $$\Delta x_t = A + \Pi x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + \sum_{s=1}^{q} \Theta_s y_{t-s} + e_t,$$

where y_t is a (m×1) vector of exogenous variables, Θ_j is a (m×m) coefficient matrices with elements $\emptyset_{jk}(i)$, and e_t is a (n×1) vector with elements e_{it}.

Weak Exogeneity Test

The weak exogeneity test determines whether or not a variable reacts to disequilibrium in the long-run. Based on the results of the test, the exogenous variables are excluded in the VAR- and VECM models and are included in the VAR-X and VECM-X models.

Equation (7) is redefined as Equation (10), replacing the error correction term (Π) by multiplication of the speed of the adjustment coefficient (α) and the long run parameter (β). We could divide Δx_t and the parameters into two parts; $\begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix}$ with dimension k_1 and k_2, A=$\begin{bmatrix}A_1\\ A_2\end{bmatrix}$,

$\alpha = \begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}$, $\phi_i = \begin{bmatrix}\phi_{1i}\\ \phi_{2i}\end{bmatrix}$, and the variance-covariance matrix $\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$.

$$(10) \qquad \Delta x_t = A + \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t$$

Then, Equation (11) could be written as:

$$(11) \qquad \begin{bmatrix}\Delta x_{1t}\\ \Delta x_{2t}\end{bmatrix} = \begin{bmatrix}A_1\\ A_2\end{bmatrix} + \begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}\beta' x_{t-1} + \sum_{i=1}^{p-1}\begin{bmatrix}\phi_{1i}\\ \phi_{2i}\end{bmatrix}\Delta x_{t-i} + \begin{bmatrix}e_{1t}\\ e_{2t}\end{bmatrix}.$$

Now, we could express the marginal model of x_{2t} as below:

$$(12) \qquad \Delta x_{2t} = A_2 + \alpha_2\beta' x_{t-1} + \sum_{i=1}^{p-1} \phi_{2i}\Delta x_{t-i} + e_{2t}$$

The hypothesis of the weakly exogenous effect of x_{2t} is $H_0: \alpha_2 = 0$. If the speed of the adjustment parameter α_2 is zero, we could conclude that x_{2t} has weak exogeneity on the other variables. This means x_{2t} does not react to a disequilibrium; also, there is no information loss even if x_{2t} is excluded.

In this research, we apply the sequential reduction method of weak exogeneity suggested by Greenslade et al. (2002). Using the standard Wald test, if a weakly exogenous variable is found in the model, we re-test the remaining variables until all weakly exogenous variables are identified (Sa-ngasoongsong et al., 2012).

Granger-Causality Test

The Granger-causality test refers to the effects of the past value of one variable on the current value of another variable. Thus, if the lags of one variable (x_{2t-1}) could improve the forecasting performance of another variable (x_{1t}), then we could say that x_{2t-1} Granger cause x_{1t}. Specifically, the equation (4) could be expressed as follows,

$$(13) \qquad \begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ x_{nt} \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \\ \cdot \\ A_{n0} \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) & \cdot & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & \cdot & A_{2n}(L) \\ \cdot & \cdot & \cdot & \cdot \\ A_{n1}(L) & A_{n2}(L) & \cdot & A_{nn}(L) \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ \cdot \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ \cdot \\ e_{nt} \end{bmatrix},$$

where A_{i0} represent the intercept parameters, polynomial $A_{ij}(L)$ are the coefficients of lagged values of variable *j* on variable *i*, and e_{it} are white-noise disturbances. If all the coefficients of $A_{ij}(L)$ are not equal to zero, we could say that variable *j* Granger cause variable *i*. The null hypothesis of the Granger-Causality test is:

$$H_0: A_{ij}(L) = 0$$

When the null hypothesis could be rejected, there exists a Granger-causality relationship. As such, the Granger-causality test is different from an exogeneity test. However, in the case of a larger VAR model (*n*>2), the Granger-causality restriction implies a weak exogeneity form. Thus, we could use the results of the Granger-causality test to confirm the results of the weak exogneity test.

Similarly, in a cointegrated process, the interpretation of the Granger-causality test is different from usual cases. Again, suppose the x_t vector in Equation (7) is $(y_t\ z_t)'$. If lagged values of Δy_{t-i} are not included in the Δz_t equation and if z_t does not respond to the discrepancy from long-run equilibrium, then we could say that $\{y_t\}$ does not Granger cause $\{z_t\}$.

Johansen's Cointegration Rank Test

Engle and Granger (1987) introduced the concept of co-integration. They consider a set of multiple nonstationary time-series variables and their long-run equilibrium. This long-run relationship between variables describes how variables adjust to deviations from equilibrium. Two conditions are necessary for cointegration. The components of vector $x_t =$

$(x_{1t}, x_{2t}, \dots, x_{nt})'$ are said to be cointegrated of order *d,b*, if first, all components of x_t are integrated of order *d*. Second, there exists a cointegrating vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ such that the linear combination $\beta' x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt}$ is integrated of order (*d*-*b*) where *b*>0. Also, the number of cointegrating vectors is called the cointegrating rank of x_t. If x_t has *n* components, *n*-1 linearly independent cointegrating vectors at most could exist. Thus, in this research, the maximum number of cointegrating vectors is 3.

Engle and Granger's (1987) method has several defects. First, it relies on a two-step estimator. Thus, step 1 errors are carried into step 2. Also, this method is not appropriate to apply with three or more variables. The estimation requires that one variable should be placed on the left-hand side, and others must be used as regressors. However, in the multivariate case, any of the variables can be placed on the left hand side. Johansen's (1988) procedure circumvents several defects of Engle and Granger's (1987) procedure. So, it could avoid two-step estimation problems and be applied to estimation and testing for the multiple co-integration vectors.

Johansen (1988) suggests two test statistics to test the null hypothesis that there are at most *r* cointegration vectors:

$$H_0 : rank(\pi) \leq r \text{ or } \pi = \alpha\beta'$$

where the speed of adjustment coefficients(α) and long-run parameter (β) are $(n \times r)$ matrices, *n* is the number of components of x_t, and *r* is rank. We could consider the term, $\beta' x_{t-1} = c$ in equation (12), as the long-run equilibrium between endogenous variables. The VECM assumes that the agents react to the disequilibrium error, $\beta' x_{t-1} - c$, and the speed of adjustment coefficient α reduce the difference between $\beta' x_{t-1}$ and c. Thus, we could consider that a large value of α implies the variable is greatly responsive to the last period's equilibrium error. Though the two rank tests share the same null hypothesis, the alternative hypotheses are different. As for the trace test, the alternative hypothesis is:

$$H_1: rank(\pi) > r$$

And the trace statistics are:

$$\lambda_{trace} = -T \sum_{i=r+1}^{p} \log(1 - \lambda_i), \tag{14}$$

where λ_i are the *p-r* smallest squared canonical correlations.

With the maximum eigenvalue test, the alternative hypothesis and test statistic are:

$$H_1: rank(\pi) \geq r + 1$$

$$\lambda_{max} = -T \log(1 - \lambda_{r+1}). \tag{15}$$

These two test results could conflict with each other. As such, the maximum eigenvalue test is considered as having the sharper alternative hypothesis. (Enders, 2003)

Forecast Encompassing Test

A preliminary comparison of the forecasting performance of the preferred consumer oriented Food and Beverage CPI forecast model is provided by the root mean square error (RMSE) and the mean absolute percentage error (MAPE). To compare the forecast of our new models with the conventional forecast provided by USDA ERS, the encompassing test is used based on Fair and Shiller (1989). We utilize their tests instead of the one proposed by Chong and Hendry (1986), which relies on error terms, because we do not know the precise model used by the USDA ERS but instead only have published reports of their forecasts over time. The equation is below:

(16) $$FCPI_t = \alpha + \lambda_1 f_{1t} + \lambda_2 f_{2t} + v_t,$$

where $FCPI_t$ is the real value of the Food and Beverages CPI, f_{1t} is the forecast value from our model, f_{2t} is the published forecast from the USDA, λ_i are the coefficients of *i* th forecast, and v_t is the error term.

If we are able to eject $H_0: \lambda_1 = 0$ and fail to reject $H_1: \lambda_2 = 0$, then it would indicate redundancy of f_{2t}. That is, the f_{1t} forecast encompasses the f_{2t} forecast. In the same vein, for switching the null and alternative hypothesis, the interpretation is in the opposite direction. Also, when both null and alternative hypotheses are rejected at the same time, it indicates that the combined (weighted) forecast with f_{1t} and f_{2t} provides a better forecast.

Results

Weak Exogeneity Test and Granger-Causality Test

Table 3-1 shows the first results of the sequential reduction method for weak exogeneity. The null hypothesis of a weak exogenous variable is rejected at the 1% level for *FCPI* and *GTI*, and the same is true for *ACPI* at the 5% level. However, we fail to reject the null hypothesis for *ICS*. For the next step, we exclude ln ICS, and then re-test the remaining variables. As Table 3-2 indicates, the null hypothesis is rejected for ln $FCPI$, ln $ACPI$, and ln GTI at the 5% level, which means these variables are endogenous. On the other hand, we can say that ln ICS does not react to disequilibrium in the long-run. Also, even if we exclude the variable in the VAR and VECM models, theoretically, there is no information loss. Thus, we exclude ln ICS in the VAR and VECM models and include ln ICS as the exogenous variable in the VAR-X and VECM-X models. In this manner, we expect that the root mean square error (RMSE) and mean absolute percentage error (MAPE) of VAR (and VECM) will be smaller than those of VAR-X (and VECM-X). These results also imply that the search based index $\ln GTI$, performs better in predicting the Food and Beverages CPI than the survey based index $\ln ICS$.

The Granger-causality test can be used to confirm the results of the weak exogeneity test. Table 2-3 indicates the results of the Granger-causality test based on the VAR and VECM models. As for the VAR, test 1 and test 3 reject the null hypothesis at the 1% significance level and test 2 does so at the 5% level, which means that group 1 variables ($\ln FCPI$, $\ln ACPI$ and $\ln GTI$) are influenced by group 2 variables (other variables except for $\ln FCPI$, $\ln ACPI$ and $\ln GTI$, respectively). On the other hand, $\ln ICS$ does not Granger cause $\ln FCPI$, $\ln ACPI$ and $\ln GTI$. Thus,$\ln ICS$ is chosen as the exogenous variable in the VAR and VAR-X models. The results of the Granger-causality test based on the VECM are similar to those based on the VAR. Test 1 and test 3, and test 2 reject the null hypothesis at the 1% level and 5% level, respectively, which is the same as the weak exogeneity test. Thus, we determine $\ln FCPI$, $\ln ACPI,$ and $\ln GTI$ are endogenous variables and $\ln ICS$ is exogenous for the VECM and VECM-X models.

Johansen's Cointegration Test

Because the variables are non-stationary over time and all have a single unit root, Johansen's cointegration rank test is conducted to determine whether a long-run equilibrium relationship exists between variables. Table 3-4 shows the results of Johansens's cointegration test. Based on both trace and maximum eigenvalue tests, we fail to reject the null hypothesis of two cointegration vectors at the 5% level. Table 2-5 indicates the long-run equilibrium relationship in the VECM model, which consists of the long run parameter β and the adjustment coefficient α with $\ln FCPI$ normalized. Two long-run relationships between three endogenous variables are:

(17) $$\ln FCPI = 1.12062 \ln ACPI + 0.05900 \ln GTI$$

(18) $$\ln FCPI = 1.35570 \ln ACPI - 0.05500 \ln GTI$$

Rolling Window Forecasting Performance Comparison

Based on the moving window and the expanding window versions of rolling windows, we evaluate the forecasting performance of the resulting models. In this research, we define the term 'moving window' to refer to the model estimates based on a fixed five years of monthly (N=60) samples of the data. Thus, we measure the first one step ahead forecast values using the first 60 observations, and for the second one step ahead forecast values, we drop the very observations and include the 61st sample. Second, 'the expanding window' refers to the model forecasts based on a total sample of the data, so the size of the window increases by one as time goes by. Initially, it is supposed that we have only five years (total 60) data and forecast the 61st values. Then, to estimate 62nd forecast values, all observations are used.

With the expanding window scheme, if structural change occurs, then the parameter estimates and forecasts would be biased and accumulated bias causes larger mean squared errors. However, reducing the number of observations in order to reduce impacts of structural change could also lead to increasing the variance of parameter estimates, which could be related to large mean squared errors (Clark and McCraken, 2009).

In practice, while the expanding window scheme is frequently used in macroeconomics literature, the moving window scheme is frequently used in financial literature. In this manner, the United States Department of Agriculture Economic Research Service (USDA ERS) also uses the expanding window scheme to forecast Food CPI. In this research, to check which scheme works better for the forecast model with consumer related index, we use both schemes to find the best consumer-oriented forecast model.

Tables 3-6 and 3-7 denote the results of assessing the predictive performance of the forecast models in both moving window and expanding window schemes using the root mean square error (RMSE) and the mean absolute percentage error (MAPE) of each forecasting model. According to Table 3-6, the VAR outperforms ARIMA-X, VAR-X, VECM and VECM-X models under the moving window scheme. Also, Table 3-7 shows that the VECM model performs better than other models under the expanding window scheme. When we compare the RMSE and MAPE of each model under two different structures, the VECM with the expanding window has smaller RMSE and MAPE than the VAR with the moving window. Though comparing the absolute values of RMSE and MAPE between two schemes could not give us a meaningful interpretation, at least we find that the expanding window scheme is more useful to apply to the consumer-oriented Food and Beverages CPI forecast model than the moving window approach.

Forecast Encompassing Test

To identify whether the consumer-oriented measurement outperforms the conventional measurement to forecast Food and Beverages CPI, we conduct an encompassing test with the suggested forecast model and reported United States Department of Agriculture Economic Research Service (USDA ERS) Food CPI forecasts. While the USDA ERS has reported the yearly Food and Beverages CPI forecasts, our estimated forecasts are monthly. To put the two forecasts on an even playing field, we convert our monthly forecasts to an annual forecast by taking an average of our models' 12 months' forecasts.

We do not know the precise model used by USDA to forecast annual CPI values, so we rely on their published forecast values. Despite knowledge of the precise models used at each point in time, Kuhns et al. (2015) describes their overall approach. Kuhns et al. (2015) indicate that for the forecast of Food CPI's subcategories, the USDA ERS uses the vertical price transmission error correction method (ECM) approach and the autoregressive moving-average approach. The selection of the methodology depends on data availability. If they can obtain the sub-categories' information of multiple stages involved in the U.S. food supply system and the food categories' data are cointegrated order *r*, then the vertical price transmission ECM methodology is applied. However, if such data limitation about a sub-categories exists, the traditional forecast model—the autoregressive moving-average approach—is used. To get the forecasts for aggregate food categories, the USDA calculates the weighted average of the forecasted subcategories' CPI.

The expanding window scheme is used to compare the performance of our estimated VECM and the USDA reported forecasts. Table 3-8 indicates that we reject the null hypothesis of $H_0\colon \lambda_1 = 0$ and fail to reject the alternative hypothesis of $H_1\colon \lambda_2 = 0$, which means that the VECM forecast using the consumer oriented variable information encompasses the USDA ERS forecast information.

Conclusions

We examine whether unconventional consumer-oriented measures improve the accuracy Food and Beverages Consumer Price Index (CPI) predictions. The exogeneity test suggests that the consumer sentiment indicator ICS does not react to disequilibrium, and thus there is no information loss even if the ICS is excluded. This result might be because the survey-based index would perform better when it is by itself rather than combined with other variables. On the other hand, we include the variable GTI, which represents consumers' interests on food prices as measured by Google internet searches, as the endogenous variable in the forecast process. Interestingly, this result supports the argument of Vosen and Schmidt (2011); the GTI outperforms the ICS in terms of forecast performance.

To access the forecast performance of competing forecast models under the moving window and expanding window scheme, we measure minimum RMSE and MAPE statistics. This preliminary comparison shows that VAR and VECM are the preferred models with the moving window and expanding window scheme, respectively. Thus, the models assuming GTI and CPI as endogenous variables best predicts the Food and Beverage CPI.

Another purpose of this research was to determine whether the consumer oriented forecast outperformed the conventional USDA ERA forecast. The encompassing test shows that the consumer oriented VECM encompasses the information contained in the USDA ERS forecast. However, this result does not mean that the USDA ERS forecasts are not valuable or inefficient, but the results suggest accuracy could be improved by including Google search data. As we discussed, these search data might have forecasting power because food prices are volatile and food is purchased frequently, which make people attentive to changes in food price.

References

Ang, A., Bekaert, G., & Wei, M. (2007). "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics* (54): 1163-1212.

Bessec, M. (2013). "Short-Term Forecasts of French GDP: A Dynamic Factor Model with Targeted Predictors." *Journal of Forecasting* 32: 500-511.

Broda, C., Leibtag, E., & Weinstein, D.E. (2009). "The Role of Prices in Measuring the Poor's Living Standards." *Journal of Economic Perspectives* 23(2): 77-97.

Brown, A., & Deaton, A. (1972). "Surveys in Applied Economics: Models of Consumer Behavior." *Economic Journal* 82(328): 1145-1236.

Carroll, C.D. (2003). "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics* 118(1): 269-298.

Carroll, C.D., Fuhrer, J.C., & Wilcox, D.W. (1994). "Does Consumer Sentiment Forecast Household Spending? If So, Why?" *American Economic Review* 84(5): 1397-1408.

Choi, H., & Varian, H. (2012). "Predicting the Present with Google Trends" *Economic Record* 88(s1): 2-9.

Chong, Y.Y., & Hendry, D.F. (1986). "Econometric Evaluation of Linear Macro-economic Models." *The Review of Economic Studies* 53(4): 671-690.

Clark, T.E., & McCracken, M.W. (2009). "Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts." *International Economic Review* 50(2): 363-395.

Dougherty, A., & Order, R.V. (1982). "Inflation, Housing Costs, and the Consumer Price Index." *American Economic Review* 72(1): 154-164.

Enders, Walter. (2003). *Applied Econometric Time Series*. Wiley Series in Probability and Statistics.

Engle, R.F., &.Granger, C.W. (1987). “Co-Integration and Error Correction Representation, Estimation, and Testing.” *Econometrica* 55(2): 251-276.

Fair, R.C., & Shiller, R.J. (1989). “The Informational Content of ex ante Forecasts.” *The Review of Economics and Statistics*: 325-331.

Fang, Y. (2003). “Forecasting Combination and Encompassing Tests.” *International Journal of Forecasting* 19: 87-94.

Fantazzini, D. (2014). “Nowcasting and Forecasting the Monthly Food Stamps Data in the US Using Online Search Data.” *PloS one* 9(11): e111894.

Giacomini, R., & Rossi, B. (2010). “Forecast Comparisons in Unstable Environments.” *Journal of Applied Econometrics* 25(4): 595-620.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). “Detecting Influenza Epidemics Using Search Engine Query Data.” *Nature* 457(7232): 1012-1014

Girardi, A., C. Gayer, & Reuter, A. (2015). “The Role of Survey Data in Nowcasting Euro Area GDP Growth.” *Journal of Forecasting* 35: 400-418

Greenlees, J.D. & McClelland, R.B. (2008). *Addressing Misconceptions about the Consumer Price Index*. U.S. Department of Labor, Bureau of Labor Statistics, Monthly Labor Review. Washington DC, 1, August.

Greenslade, J.V., Hall, S.G., & Henry, S. B. (2002). “On The Identification of Cointegrated Systems in Small Samples: A Modelling Strategy with an Application to UK Wages and Prices.” *Journal of Economic Dynamics and Control* 26(9): 1517–1537.

Johansen, S. (1988). “Statistical Analysis of Cointegration Vectors” *Journal of Economic Dynamics and Control* 12(2): 231-254

Kuhns, A., Volpe, R., Leibtag, E., & Roeger, E. (2015). *How the USDA Forecasts Retail Food Price Inflation.* Washington DC: U.S. Department of Agriculture, Economic Research Service, Technical Bulletin No TB-1940, May.

Joutz, F.L. (1997). "Forecasting CPI Food Prices: An Assessment." *American Journal of Agricultural Economics* 79(5): 1681-1685.

Pearce, D.K. (1979). "Comparing Survey and Rational Measures of Expected Inflation: Forecast Performance and Interest Rate Effects." *Journal of Money* 11(4): 447-456.

Sa-ngasoongsong, A., Bukkapatnam, S.T., Kim, J., Iyer, P.S., & Suresh, R.P. (2012). "Multi-step Sales Forecasting in Automotive Industry Based on Structural Relationship Identification." *International Journal of Production Economics* 140: 875-887.

Sims, C.A. (1980). "Macroeconomics and Reality." *Econometrica* 48(1): 1-48.

Sims, C.A., Stock, J.H., & Watson, M.W. (1990). "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica* 58(1)*:* 113-144.

Souleles, S. N. (2004). "Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Survey." *Journal of Money* 36(1): 39-72.

Stock, J.H., & Watson, M.W. (1999). "Forecasting Inflation." *Journal of Monetary Economics* 44: 293-335.

Souchoy, T. (2009). *Query indicies and a 2008 downturn: Israeli data.* Bank of Israel. Research Department, http://www.bankisrael.gov.il/deptdata/mehkar/papers/dp0906e.pdf

Swallow, Y. C., & Labbe, F. (2013). "Nowcasting with Google Trends in an Emerging Market." *Journal of Forecasting* 32(4): 289-298.

Swanson, N.R. (1998). "Money and Output Viewed Through a Rolling Window." *Journal of Monetary Economics* 41(3): 455-474.

Toda H Y., & Taku, Y. (1995). “Statistical Inference in Vector Autoregressions with Possibly Integrated Processes.” *Journal of Econometrics* 66(12): 225–250.

Tseng, F.M., Tzeng, G.H., Yu, H.C., & Yuan, B.J. (2001). “Fuzzy ARIMA Model for Forecasting the Foreign Exchange Market.” *Fuzzy Sets and Systems* 118: 9-19.

Van Der Voort, M., Dougherty, M., & Watson, S. (1996). “Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow.” *Transportation Research Part C* 4(5): 307-318.

Vosen, S., & Schmidt, T. (2011). “Forecasting Private Consumption: Survey-Based Index vs. Google Trends.” *Journal of Forecasting* 30(6): 565-578.

Wilcox, J. (2007). “Forecasting Components of Consumption with Components of Consumer Sentiment.” *Business Economics* 42(4): 22-32.

Zivot, E., & Wang, J. (2007). *Modeling Financial Time Series with S-Plus®,* Vol. 191. Springer Science & Business Media

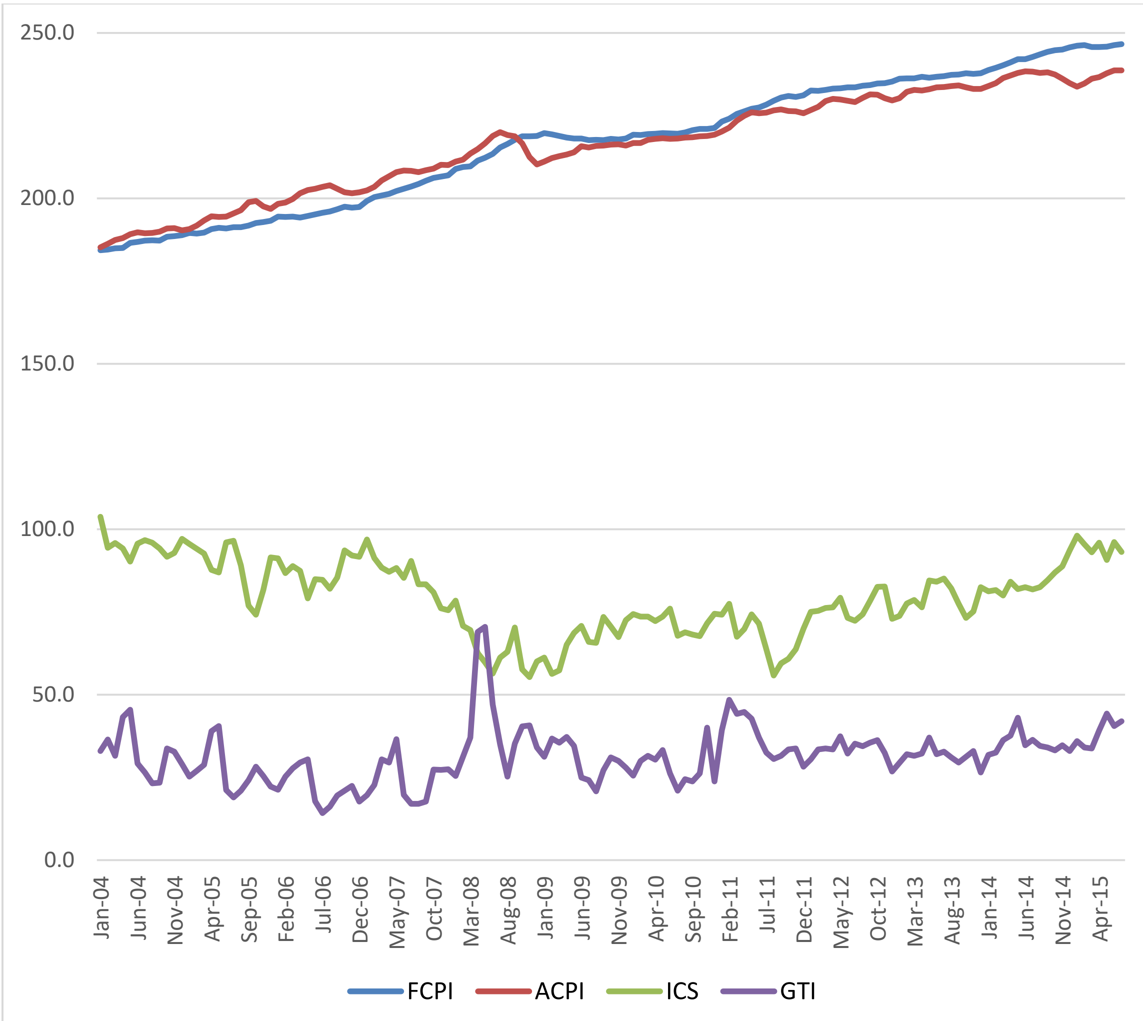


Figure 3-1. Plot of Food and Beverages CPI (FCPI), All Items CPI (ACPI), Google Trend Index (GTI) and Index of Consumer Sentiment (ICS) between 2004 and 2015

Table 3-1. The Results of Weak Exogeneity Test (All Variables)

Variable	χ^2	$\Pr > \chi^2$
$\ln FCPI$	19.86	0.0002***
$\ln ACPI$	7.85	0.0491**
$\ln GTI$	46.59	<.0001***
$\ln ICS$	3.75	0.2902

The last column entry is the p-value of the null hypothesis of a weak exogenous variable. The asterisk *, double **, and triple *** indicate the null hypothesis can be rejected at the 0.10, 0.05, and 0.01 levels, respectively.

Table 3-2. The Results of Weak Exogeneity Test (Re-Test)

Variable	χ^2	$\Pr > \chi^2$
$\ln FCPI$	18.43	<.0001***
$\ln ACPI$	6.41	0.0406**
$\ln GTI$	44.33	<.0001***

Based on Table 3-1, we re-test the remaining variables. The last column entry is the p-value of the null hypothesis of a weak exogenous variable. The asterisk *, double **, and triple *** indicate the null hypothesis can be rejected at the 0.10, 0.05, and 0.01 levels, respectively.

Table 3-3. The results of Granger-causality Test

Tests	VAR			VECM		
	Optimal Lag	χ^2	$\Pr > \chi^2$	Optimal Lag	χ^2	$\Pr > \chi^2$
1	2	22.76	0.0009***	2	24.36	0.0004***
2	2	10.96	0.0897*	2	11.73	0.0683*
3	2	28.92	<.0001***	2	30.95	<.0001***
4	2	6.19	0.4025	2	6.62	0.3571

The asterisk *, double **, and triple *** indicate the null hypothesis can be rejected at the 0.10, 0.05, and 0.01 levels, respectively. Test 1: Group 1 is $\ln FCPI$ and Group 2 is $\ln ACPI$, $\ln GTI$, $\ln ICS$.
Test 2: Group 1 is $\ln ACPI$ and Group 2 is $\ln FCPI$, $\ln GTI$, $\ln ICS$.
Test 3: Group 1 is $\ln GTI$ and Group 2 is $\ln FCPI$, $\ln ACPI$, $\ln ICS$.
Test 4: Group 1 is $\ln ICS$ and Group 2 is $\ln FCPI$, $\ln GTI$, $\ln ACPI$.

Table 3-4. Johansen's Cointegration Rank Tests

Trace Test			
$H_0: Rank = r$	$H_0: Rank > r$	Trace Statistics	5% Critical Value
0	0	99.318	29.38
1	1	20.088	15.34
2	2	2.648	3.84
Maximum Eigenvalue Test			
$H_0: Rank = r$	$H_0: Rank = r + 1$	Max Statistics	5% Critical Value
0	1	79.230	20.97
1	2	17.441	14.07
2	3	2.648	3.76

Based on Table 1 and 2, ln $FCPI$, ln $ACPI$, and ln GTI are used for Johansen's cointegration rank tests.

Table 3-5. Long-Run Parameter β Estimates and Adjustment Coefficient α Estimates (Rank=2)

	Long-run β		Adjustment coefficient α	
Variable	1	2	1	2
$\ln FCPI$	1.000	1.000	-0.048	-0.021
$\ln ACPI$	-1.121	-1.356	0.039	0.027
$\ln GTI$	-0.059	0.055	6.190	-2.024

Table 3-6. 1-Step Ahead Food and Beverage CPI Forecasting Comparison Using RMSE and MAPE by Moving Window Scheme

Models	RMSE	MAPE
ARIMA-X	0.00117	0.01716
VAR	0.00086	0.01159
VAR-X	0.00097	0.01283
VECM	0.00090	0.01249
VECM-X	0.00110	0.01442

Table 3-7. 1-Step Ahead Food and Beverage CPI Forecasting Comparison Using RMSE and MAPE by Expanding Window Scheme

Models	RMSE	MAPE
ARIMA-X	0.00089	0.01281
VAR	0.00080	0.01088
VAR-X	0.00090	0.01183
VECM	0.00075	0.01060
VECM-X	0.00086	0.01154

Table 3-8. Encompassing Test

Models	t-value	Pr >t
USDA model	2.01	0.1002
VECM(2)	15.26	<.0001***

The last column entry is the p-value of the null hypothesis that $H_0: \lambda_1 = 0$ and $H_1: \lambda_2 = 0$, respectively. The asterisk *, double **, and triple *** indicate the null hypothesis can be rejected at the 0.10, 0.05, and 0.01 levels, respectively.

Appendix

Unit root test

The Augmented Dickey-Fuller unit root test identifies whether the variables are stationary over time. The general to specific methodology (t-test) and measurement of model selection—Akaike Information Criteria (AIC) and Schwarz Bayesian Criterion (SBC)—are used to select the optimal lag for the unit root test. When the results are different, we choose the lag which is selected at least by two criteria. As for the *lnFCPI* in level, *lnFCPI* in difference, *lnAllCPI* in difference and *lnICS* in difference, the result of general to specific test are consistent with that of AIC and SBC. On the other hand, *lnAllCPI* in level, *lnGTI* in level, *lnICS* in level, and *lnGTI* in difference do not have the same results between criteria. For the *lnAllCPI* in level, the second lag is selected as the optimal lag by t-test and SBC. And the fifth, third, and sixth lag are chosen by t-test and AIC for *lnGTI* in level, *lnICS* in level, and *lnGTI* in differences, respectively.

Table A3-2 presents the Augmented Dickey-Fuller unit root test results. We fail to reject the null hypothesis of a unit root for the variables in levels at the 1% significance level, and the null hypotheses of a unit root for the first differenced variables are rejected at 5% level, which means that the variables taking the first difference do not have unit roots. Thus, we obtain stationary variables using first differences.

Table A3-1 Information Criteria for Selection of Optimal Lag for Unit Root Test

Variables	Lag	AIC	SBC	Variables	Lag	AIC	SBC
$\log(FoodCPI)$	**6**	**-1273.07**	**-1252.53**	$\Delta\log(FoodCPI)$	6	-1279.51	-1259.02
	5	-1264.18	-1246.57		5	-1281.51	-1263.95
	4	-1263.91	-1249.24		4	-1282.00	-1267.37
	3	-1233.99	-1222.25		**3**	**-1282.92**	**-1271.21**
	2	-1222.62	-1213.82		2	-1274.70	-1265.92
	1	-1189.50	-1183.63		1	-1276.10	-1270.25
$\log(AllCPI)$	6	-1137.55	-1117.00	$\Delta\log(AllCPI)$	6	-1155.08	-1134.59
	5	-1140.14	-1122.54		5	-1156.75	-1139.18
	4	-1143.14	-1128.46		4	-1155.28	-1140.64
	3	**-1145.49**	-1133.75		3	-1157.09	-1145.38
	2	-1145.10	**-1136.30**		**2**	**-1158.38**	**-1149.60**
	1	-1089.20	-1083.33		1	-1151.14	-1145.28
$\log(GTI)$	6	-90.48	-69.94	$\Delta\log(GTI)$	**6**	**-90.20**	-69.71
	5	**-92.12**	-74.51		5	-84.84	-67.27
	4	-80.98	-66.31		4	-86.83	**-72.20**
	3	-80.54	-68.80		3	-69.00	-57.29
	2	-82.53	-73.73		2	-61.69	-52.90
	1	-80.67	**-74.80**		1	-60.13	-54.28
$\log(ICS)$	6	-394.03	-373.49	$\Delta\log(ICS)$	6	-390.08	-369.59
	5	-394.75	-377.15		5	-391.69	-374.13
	4	-396.74	-382.07		4	-392.30	-377.66
	3	**-396.96**	-385.23		3	-394.27	-382.56
	2	-391.77	-382.96		**2**	**-394.37**	**-385.59**
	1	-393.66	**-387.79**		1	-388.71	-382.86

Each bold in "Lag" column indicates the significant lag by the general to specific methodology (t-test). Each bold in both "AIC" and "BIC" columns indicate the lag has the smallest values of each measurement of model selection.

Table A3-2 Augmented Dickey-Fuller Unit Root Tests

Variables	Optimal lags	Zero mean		Single mean		Trend	
		τ_μ	$Pr < \tau_\mu$	τ_μ	$Pr < \tau_\mu$	τ_μ	$Pr < \tau_\mu$
$\log(FoodCPI)$	6	2.214	0.994	-1.011	0.747	-2.208	0.481
$\log(AllCPI)$	2	3.008	0.999	-1.549	0.506	-2.712	0.234
$\log(GTI)$	5	0.062	0.702	-2.439	0.133	-3.286	0.073
$\log(ICS)$	3	-0.103	0.647	-1.807	0.376	-1.518	0.819
$\Delta\log(FoodCPI)$	3	-2.348	0.019	-3.644	0.006	-3.696	0.026
$\Delta\log(AllCPI)$	2	-5.604	<.0001	-6.585	<.0001	-6.707	<.0001
$\Delta\log(GTI)$	6	-4.967	<.0001	-4.951	0.0001	-4.941	0.0005
$\Delta\log(ICS)$	2	-8.627	<.0001	-8.595	<.0001	-8.717	<.0001

VITA

Jisung Jo

Candidate for the Degree of

Doctor of Philosophy

Thesis: VALUE OF PARSIMONIOUS NUTRITIONAL INFORMATION, PREDICTING FOOD PRICE, AND CONSUMER-ORIENTED FOODS CLUSTER

Major Field: Agricultural Economics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Agricultural Economics at Oklahoma State University, Stillwater, Oklahoma in December, 2016

Completed the requirements for the Master of Science/Arts in Agricultural Economics at Seoul National University, Seoul, Korea in 2013.

Completed the requirements for the Bachelor of Science in Agricultural Economics at Pusan National University, Busan, Korea in 2011.

Experience:

Graduate Research Associate, 2014-2016
Department of Agricultural Economics, Oklahoma State University

Graduate Research Assistant, 2011-2013
Department of Agricultural Economics, Seoul National University