Estimating Price Effect on Consumer Shopping Across Quality-Differentiated Store Formats*

Ming-Feng Hsieh**

Price is a key to consumer's shopping decisions. Therefore, understanding and estimating price effect on consumer shopping behavior is constantly of concern in marketing analysis and economic research. This empirical study employed the A.C. Nielsen Homescan Data to examine consumers' grocery shopping across quality-differentiated store formats and the impacts of prices on consumer store format choice decision in a non-coastal U.S. City during 2005-2008. Results from the Mixed Multinomial Logit regression show that the price effect on consumer store format choice varies among formats, and the variation in price effect is smaller among the high-end format's shoppers than among the general formats' shoppers. In addition, the results show that the price effect varies among consumers of different income levels, and the richer is less price sensitive compared to the poorer. The method and results from this empirical study provide useful insights for understanding the effect of prices on choosing where to shop in a market with preference heterogeneous consumers and quality differentiated retailers.

***Keywords**: Price Effect, Store Format Choice, Mixed Multinomial Logit Model, Supermarket, Food Retailing*

---

* I thank Kyle W. Stiegert, Jean-Paul Chavas, Timothy J. Richards, Paul D. Mitchell, and Brian W. Gould for helpful comments and suggestions. I am also grateful to Dr. Ephraim Leibtag and Dr. Biing-Hwang Lin at Economic Research Service of the U.S. Department of Agriculture, and the A.C. Nielsen Company for supplying key data for this study. I thank two referees for helpful comments and suggestions.

** Assistant Professor, Department of Bio-Industry Communication and Development, National Taiwan University, Taiwan.

農業經濟叢刊 (Taiwanese Agricultural Economic Review)，18:1 (2012)，1-30。
臺灣農村經濟學會出版

I. Introduction

Pricing and promotion strategies have long been considered the most frequently adopted methods for retailers to increase store traffic. Especially in food markets, consumers are constantly attracted to promotional discounts and are sensitive to price increases probably due to weather factors. We have seen the success of emphasizing on pricing strategies from the operations of price clubs or supercenters, such as Wal-Mart and Target (Basker, 2005; Basker and Noel, 2007; Franklin, 2001; Stiegert and Sharkey, 2007). However, it is also commonly accepted that there are many other important factors that affect consumers' shopping decisions. Research has shown that retailers' assortment, service quality, location, and consumers' demographics, along with prices, affect shopping behavior. Many found that those non-price factors are more important determinants than price on consumer store choice and purchase decisions (Carpenter and Moore, 2006; Fox, Montgomery, and Lodish, 2004).

These arguments raise an important question in retail strategy and marketing research: To what extent can a store's pricing and promotional activities be used to increase the store's traffic and how will the price effects be dependent of those non-price marketing strategies? The linkages between price and non-price factors are especially important in estimating price effect on consumer shopping across quality-differentiated store formats, because consumers are likely to have different perceptions on prices when stores' products and services vary.

To address this question, I take the perspective of a shopper and analyze the underlying factors that affect store format choice. With the emphasis on the concept of value, which incorporating both quality and price, I analyze how the price effect varies among the households with heterogeneous willingness-to-pay and quality perceptions and among the store formats with differential quality settings. This empirical analysis uses household-side scanner panel data, which collect the households' food and grocery purchases from all formats of outlets. The three store formats for this study include "value-oriented retailers" featuring low-pricing policy, traditional "supermarkets" representing a middle format, and "high-end" specialty stores providing high priced organic and gourmet foods. The household-level scanner data contain demographic information and the residential location, which enable us to analyze how certain non-price factors affect consumer shopping decisions.

Previous empirical research on store choice and sales tends to estimate the impacts of the price and the non-price factors independently (Bell, Ho, and Tang, 1998; Bell and Lattin, 1998; Bhatnagar and Ratchford, 2004; Ho, Tang, and Bell, 1998). The heterogeneity in consumers' preferences has been left out from these studies. Till the recent development in discrete choice models with unobserved heterogeneity, Briesch, Chintagunta, and Matzkin (2010) provide the estimation on consumer response to cost and distance taking into accounts of their interactions and nonlinear effects in a model of store format choice. The present article adopts a similar approach but adds the income factor to explain the variation in price effects among households. As demonstrated in the author's another theoretical work, Hsieh (2012), income is likely to capture the household's willingness to pay and quality perception. So it can be used to control for the heterogeneity in household's price sensitivity in store format choice decision.

The remainder of this paper is organized as follows. First, I construct an econometric model that quantifies the impacts of economic and socio-demographic factors on the probability of a household choosing a specific store format for food and grocery purchases. I then discuss the methods used for estimating the price effect on shopping location choice among quality differentiated store formats. The data and descriptive statistics are discussed next. I then present and discuss the regression results. Next, I discuss how the marginal effects of price vary among various groups of income. The paper concludes with a summary of findings and remarks.

II. Consumer Shopping Across Retail Formats

In the present study, I build an empirical analysis using the theoretical framework in Hsieh (2012) to model consumers' grocery shopping decision. In such a framework, consumers, who are heterogeneous in preference, choose where to shop among three quality and price differentiated retail formats. Following the setup, three differential retail formats are considered: low-end value-oriented retailers (L), the middle format: traditional supermarkets (M) and high-end specialty stores (H). Consumers make two store format choices: one for the targeted product category – food (available at all three formats) and the other for the remaining items on a shopping list (available only from the general formats: L and M). Consumers are considered to be heterogeneous in both fixed and variable willingness to pay (WTP) for product quality. As illustrated in Figure 1, heterogeneous households may be categorized into big-basket, small-basket or no-purchase shoppers based on the size of their shopping basket derived from their needs and retailers' strategies. Furthermore, we can divide those big-basket shoppers into the two groups: one-stop or cross shopping based on whether the households shopped in one store format or two. The model assumes that shoppers face a fixed cost for each shopping trip; hence, additional cost would occur for those who do cross shopping if compared to the one-stop shoppers.

By identifying the matching between consumers' WTPs and their optimal store format choice combinations, the model provides a foundation to understand consumer responses to retailers' quality setting and pricing. For the former, with the model setup, an increase in a retailer's quality setting enhances consumers' overall evaluation for the store and encourages more store visits. Such an impact of quality change on consumer's fixed satisfaction from store's service is called the *store image effect*. In addition, a quality improvement would affect consumers' perception of the store's value, which in turns influences their purchase decisions including store format choice as well as the amount of quantity purchased or the expenses, so called the *value effect*.

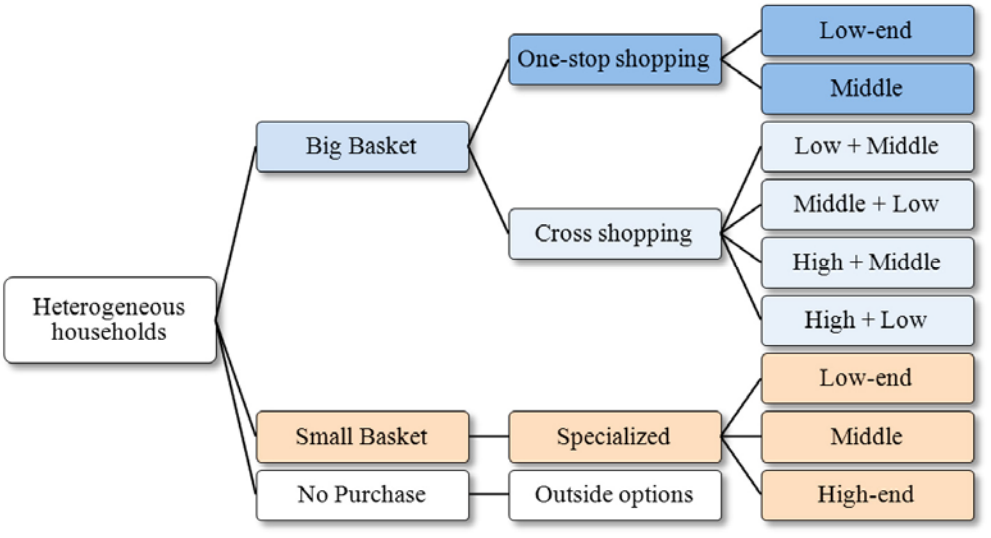


Figure 1. Households' Purchase & Store Format Choice Decision Making

Data source: This study.

These two dimensional impacts of quality setting on consumer shopping across store formats play a key role in the findings of this framework. For the latter, the model shows that *the price effects are dependent of store format's quality, consumer's income and unobserved preference heterogeneity*. Particularly, the model leads to the following useful and testable implications for the impacts of price on consumer shopping across retail formats.

First, *the price effects on store format choice vary among households of different types and among formats* (Hypothesis, H1). Consumers of different types would be attracted to the formats that offer a matching combination of price and quality positioning to their tastes. Thus, price sensitivity is expected to vary across store formats and among households of different types, likely reflected by income, preference for organic consumption and demographic characteristics. In addition, variation in consumer unobservable preference heterogeneity also directs price sensitivity to vary among consumers.

Second, *variation in price sensitivity is smaller among the high-end format's shoppers than among the general formats' shoppers* (Hypothesis, H2). As demonstrated by Hsieh (2012)'s results, high-end specialty format's shoppers are with high fixed and high variable WTPs while the other two formats' shoppers are with diversified WTPs. Given the price sensitivity is closely tied with variable WTP, it is expected that variation in price sensitivity follows the similar pattern of variation in consumers' WTPs among the shoppers of differential formats.

Finally, since consumer's store format choice is a choice decision concerning multiple products; hence, the inter-category connection, especially cross-price effects, will have impacts on store format choice. As a result, we may observe certain format combinations that are appealing to a group of consumers but not others. Shopping costs revealed by the distance between retailer and consumer will help explain store format choice. The households, who travel longer distance, shop less frequently, or have less access to a certain format, will likely bear higher travel costs; thus, they will prefer formats with broader assortments, reflecting one-stop shopping behavior. These predictions that have commonly been established in literature will too be considered in my empirical estimation.

III. Model Specification and Measures

Consumer's shopping decision on choosing where to shop, is a discrete choice problem based on consumers' utility maximization among available alternatives. That is, household *h* maximizes the following indirect utility associated with purchasing foods at a specific retail format *f* on trip *t*:

$$\begin{aligned} V_{hft} &= \hat{V}_{hft} + \varepsilon_{hft} \\ &= A_{hft}\alpha_h + X_{hft}\beta_h + \varepsilon_{hft} \end{aligned} \tag{1}$$

where $A_{hft}\alpha_h$ is the fixed component of deterministic utility, $X_{hft}\beta_h$ represents the variable component of deterministic utility which is dependent of the household's shopping list for the specific trip, and ε_{hft} is the error term being assumed with independence of irrelevant alternatives (IIA), i.e. the odds of preferring one choice over another do not depend on the presence or absence of other "irrelevant" alternatives. This specification takes into account the repeated choices by each individual decision maker by treating the coefficients that enter utility as varying over households but being constant over choice situation for each household.

Following the theoretical framework, the set of explanatory variables affecting the fixed component of utility (*A*) includes: 1) *income*, 2) *shopper types* represented by a shopping preference for organic foods and a discount use rate in this study, 3) *individual household characteristics* including educational and marital status, household size and family compositions, 4) *shopping costs and convenience* reflected by store locations, average days between trips and network sizes of store formats, and 5) *quality index* represented by the percentage of organic products to total products sold in each format.

The set of factors influential to the variable component of utility (*X*) contains 1) *prices* represented by price indexes and 2) *shopping basket* size and components. To capture the preference heterogeneity, I include an interaction terms of income and price, i.e. *income*price* and $income^2 * price$, as well as random price coefficients to reflect the variation in price sensitivity among different income groups of households.

It is worth noting that I do not include the interaction terms of prices and quality indexes to represent the value variable in the present estimation due to the multi-linearity issue with such a variable and the main term of price. However, as suggested by the theory, the differences in quality perception among consumers will cause differential price sensitivities in consumer shopping across retail formats. This calls for the Mixed Logit Models for estimation of random preference variation.

3.1 A Mixed Logit Model with Repeated Choices

The store format choice decision based on maximization of utility (1) fits the use of a Mixed Logit Model with repeated choices, which has been well established in the literature, such as Train (2003). The motivation for the Mixed Logit Model arises from the limitations of the standard Logit Model. Particularly, the Mixed Logit Model solves the shortcomings of the standard Logit Model by allowing for 1) random taste variation, 2) unrestricted substitution patterns, and 3) correlation in unobserved factors over time (Train, 2003). Therefore, the Mixed Logit Models especially outscore the standard Logit Models for estimations using panel data, like in the present study. In addition, unlike the Probit Models which are limited to the normal distribution, the Mixed Logit Models can also utilize any distribution for the random coefficients.

For estimation, I follow the simulation approach of Train (2003) and utilize the STATA command: MIXLOGIT written by Hole (2004).[1] The approach considers a sequence of alternatives (store formats), one for each trip, $f = f_1, \ldots, f_T$. The probability the the household makes this sequence of choices is the product of Logit formulas:

$$L_{hf}\left(\alpha,\beta\right)=\prod_{t=1}^{T}\left[\frac{e^{A_{hft}\alpha_h+X_{hft}\beta_h}}{\sum_j e^{A_{hfj}\alpha_h+X_{hfj}\beta_h}}\right], \tag{2}$$

where it is assumed that the ε_{hft}'s are independent over time. The unconditional probability is the integral of this product over all values of $\left(\alpha,\beta\right)$:

$$P_{hf}=\int\left(\int L_{hf}\left(\alpha,\beta\right)f\left(\beta\right)d\beta\right)f\left(\alpha\right)d\alpha\,. \tag{3}$$

The estimation of the probability for the sequence of choices consists of a great deal of computations and time due to the consideration for repeated choices, requiring

T times as many draws compared to the case where coefficients being constant over time for each household. Due to the computational issues with a fairly large size of analyzed sample facing in this study, I allow random parameters for only the quality and price variables to obtain converged estimation results. The specification of the deterministic component of latent utility adopted for estimation is:

$$\begin{aligned}\hat{V}_{hft} = {} & \alpha_1 \text{ income}_h + \sum_f \alpha_{2f} \text{ loyalty}_{hf} + \alpha_3 \text{ organic-shopper}_h + \alpha_4 \text{ \%discount}_h \\ & + \sum_j \alpha_{5j} \text{ demographic}_h^j + \alpha_6 \text{ days-between-trips}_{ht} + \alpha_7 \text{ distance}_{hft} \\ & + \sum_f \alpha_{8f} \text{ n(store)}_{hft} + \sum_f \alpha_{9fh} \text{ quality}_{hft} + \sum_f \beta_{1fh} \text{ price}_{hft} \\ & + \sum_f \beta_{2f} \text{ income}_h * \text{price}_{hft} + \sum_f \beta_{3f} \text{ income}_h^2 * \text{price}_{hft} \\ & + \beta_4 \text{ total-spending}_{hft} + \sum_k \beta_{5k} \text{ \% spending-in-product-category-}k_{hft} \end{aligned} \quad (4)$$

where α_{9fh} and β_{1fh} are the random parameters. The *demographic* variables include household size, education, marital status, children and age (elderly), and *product category* variables include dry grocery, fresh produce and organics.

3.2 Variables and Measures

The measurements of the variables, including shopper types, price index, basket size and composition, shopping cost and convenience, and demographics, are described as follows.

3.2.1 Shopper Types

To identify shopper types, I adopt the following three measures: format loyalty, discount shoppers, and organic shoppers. First, the format-specific loyalty for a household is represented by the percentage of trips that the household made to the format during the initialization period. This loyalty index, based on Bell, Ho and Tang's formula (p. 297), reveals the shopper's preference toward a specific format due to probably the familiarity about the store layout, the general prices and assortments, and the convenience and quality of service, based on his/her past shopping experience.

Second, I use household discount-use rate, defined as the ratio of the number of items purchased with coupons or at discounted prices to the total number of items in a shopping trip, to capture their preference between promotional pricing (HiLo) and everyday low pricing (EDLP). These rates are calculated from the household purchase information during the initialization period. It is expected that a household with a high discount-use rate would prefer the format in which stores/chains use HiLo pricing instead of EDLP.

Third, I use the household purchase information during the initialization period to identify whether the householder is a frequent organic shopper or not, depending on whether the percentage of organic consumption is more than 10% of the total spending. Since organic shoppers are likely to consist of high-income households discount lovers, small households, college educated householders, families with preschool children, but not families with school-age children and the elderly households, it is expected that the likelihood for frequent organic shoppers to choose high-end format stores is relatively higher than the one for non-organic shoppers.

3.2.2 Shopping Cost and Convenience

As suggested in my theoretical model as well as in store choice literature, shopping cost and convenience are important to consumer decision making in store format choice. In this research, I use “days between trips” and the average “distance” between the stores of certain format and the household’s home to measure the shopping costs. It is expected that the more days between trips reflects the higher shopping cost that the consumer may encounter for each individual trip. The longer distance between the store and household’s home is, the higher shopping cost for the household will be.

Regarding to the measure of “distance”, there is no detailed location information for both stores and households except the zip code for store and the census track code for household in the data available to us. Thus, the distance measure between a specific pair of store and household is calculated by the distance between the center of the area with the zip code of the store and the center of the area with the census track code of the household. Although this is less than ideal and can potentially cause measurement error, it is likely to yield better outcomes with less bias compared to the case without a distance measure, where the estimation suffers from the omitted variable bias.

To better represent the location impacts, I use the number of stores for each format that are located with 25 miles from the household’s home to capture the location convenience of the format, called the network effects. I expect that the more number of stores of the format is expected to have positive impacts on the store format choice for the format.

3.2.3 Demographics

The demographics available to this study include household size, educational and marital status of the householders, and the family compositions – the ages and numbers of children and the ages of the householders. Selecting the major group of consumers (college educated, married, no young children and non-elderly) as the default group, I examine the impacts of demographic characteristics differing from the default on the decision of store format choice.

3.2.4 Quality Index

Measuring store format quality represents a challenge for this study due to the limitation of this sort of data. Past researcher employed measures such as whether or not prepared food service is available, the number of product varieties, product categories, check-out lines etc, to proxy a store’s quality (e.g. Baker, Grewal, and Parasuraman, 1994; Dodds, Monroe, and Grewal, 1991; Gotlieb, Grewal, and Brown, 1994; Parasuraman and Grewal, 2000; Voss, Parasuraman, and Grewal, 1998). In this study, I adopt a new approach to measure a store's quality: the share of the store's sales coming from organic food products. The organic label covers a wide range of products, it reflects rigorous supply management on the part of food processors, and farmers must meet strict guidelines regarding the production of organic commodities. As a result, organic food is perceived as an indicator of food quality leading to perceptions by consumers about the quality of the store's product offerings (Brunsø, Fjord, and Grunert, 2002). Consequentially, I hypothesize a store's new introduction or increase in the supply rate of organic foods as an increase in the store's quality. Specifically, a store format's quality index in this study is represented by the total number of organic products sold (*org*). The log ratio of organic supply, defined as the log of ratio of numbers of organic products sold in the two formats: $\ln(org_f/org_M)$, is used to capture the impact of quality in my estimation.

3.2.5 Price Index

Applying Laspeyres index approach to generate the format-specific price index, I first select a comparable basket of items available for all three formats. Table 1 lists the base basket containing 30 most frequently purchased product categories: After the base basket is constructed, I then calculate the average household consumption pattern for the selected product categories in the basket from the initialization sample. Using these base quantities together with the format-specific category price indexes, we can compute the cost of the base basket at each format. Specifically, the cost measure can be written as:

$$CI_f^t = \sum_{j=1}^{J} p_{f,j}^t * x_{\bar{h}j}^{t_i} \tag{5}$$

where $p_{f,j}^t$ is the price index of category *j* at format *f* in the estimation month *t* and $x_{\bar{h}j}^{t_i}$ is the quantity that average household $\bar{h}$ purchases in category *j* in the initialization period t_i.

Table 1. The Base Basket for Format-specific Price Index

1. Dairy-Milk-Refrigerated	2. Bakery - Fresh Bread
3. Cereal - Ready To Eat	4. Soft Drinks - Carbonated
5. Yogurt-Refrigerated	6. Fruit
7. Soup-Canned	8. Cookies
9. Vegetables	10. Eggs-Fresh
11. Precut Fresh Salad Mix	12. Candy-Chocolate
13. Fruit Drinks-Other Container	14. Water-Bottled
15. Beef	16. Snacks - Tortilla Chips
17. Fresh Carrots	18. Fresh Strawberries
19. Fresh Fruit-Remaining	20. Rice - Mixes
21. Fruit Juice - Apple	22. Fruit Juice-Remaining
23. Prepared Foods	24. Yogurt-Refrigerated-Shakes & Drinks
25. Frozen Fruits	26. Vegetable Juice And Drink Remaining
27. Meat Products-Imitation & Additives	28. Fish
29. Whipping Cream	30. Seafood-Shellfish.

Date source: A. C. Nielsen Homescan data and arranged by this study.

With the cost measures for the formats, I use the cost for purchasing the base basket at supermarkets (format M) in the first period as the base cost to calculate the ratio of every cost measure to this base cost. This yields the format-specific price index:

$$PI_f^t = CI_f^t \big/ CI_M^t \ . \tag{6}$$

3.2.6 Basket Size and Composition

For measuring the basket size of a shopping list, there are a “size” approach, using total number of items purchased, and a “value” approach, using total spending, in literature. Following Bell, Ho, and Tang (1998), I adopt the value approach, i.e. use “total spending,” defined by the total transaction amount recorded, to measure the shopping basket size for each individual shopping trip. Unlike the size approach, this basket size measure captures the household’s monetary cost of shopping, consisting of both the number of items and the monetary value for each item purchased during the shopping trip. In addition, I use the average shares of dry grocery, fresh produce and organic foods in the total consumption during the initialization period to measure how consumers' consumption preference pattern affects their choice of store format. Instead of directly using the basket composition of the current trip, this alternative measure can reduce the bias due to endogeneity.

IV. The Data

I use a multi-outlet panel data set－Homescan by A. C. Nielsen for a non-coastal U.S. city that covers a 208-week period between December 26, 2004 (hereafter January, 2005) and December 27, 2008. The data panel includes a representative of households in 52 market areas and 9 remaining (rural) areas in the 48 continental states of the US. Panelists report their purchases by scanning either the Uniform Product Code (UPC) or a designated code for random weight products of all their purchases from grocery stores or other retail outlets.

The data set contains purchase/sales information on eight product departments (dry goods, frozen, dairy, deli, meat, fresh produce, non-food, and alcoholic beverages) and over 600 product categories of food and non-food items sold in grocery stores or other retail outlets. The households report weekly purchase data including price, quantity, promotional information, and product characteristics. One of the product characteristics contained in the data is the identifier for organic products. For UPC-coded products, organic products can be identified by the presence of the USDA organic seal or with organic-claim codes created by A. C. Nielsen. In addition, the data set contains store-level information including store location, store size (as represented by an estimate of annualized value of sales), store formats (as described below), and demographics of store users. For identifying store formats, the retail outlets included in the data set are categorized into eight groups: 1) grocery stores, 2) drug stores, 3) mass merchandisers, 4) supercenters, 5) clubs, 6) convenience stores, 7) health food stores, and 8) all other.

The household demographic information is collected for each panelist, including household size, income, age, employment, education, marital status, race, type and location of residence, and selected household equipment characteristics (e.g. kitchen appliances, TV items, internet accessibility). These individual household demographic characteristics will be aggregated by store based on the purchase/sale information to generate the store-level demographics of store users.

This is a useful data set to study the linkages of organic consumption and store format choice in the US grocery sector. It will allow for an analysis of the role of major retailers such as Wal-Mart and Whole Foods that traditionally do not cooperate with scanner data collection companies, since the data are recorded from the consumers' end instead of by the retailers. The rich information of household purchasing pattern, especially with the organic claim, allows me to study the present research that cannot be addressed using other forms of data.

In the database available and used for my analysis, households are included only if they participated in at least 10 of the 12 months during the year; these households are referred to as the "static" panel. Only a subset of this static panel, called the Fresh Foods panel, records purchases of random-weight foods without UPCs. During 2005-06, the full static panel consists of approximately 40,000 households per year, with 8,000 recording both UPC and random weight items and 32,000 recording only UPC items. In 2007, Nielsen replaced the Fresh Foods Panel with the Total Shopper View, which no longer contains details on the random weight items. Specifically, item characteristics and the quantity purchased are no longer recorded. However, the sample size increased significantly to 22,000 households for those households reporting both UPC and non-UPC items, and a total sample of 60,000 households recording just UPC purchases. Due to the inconsistency on the coverage of random weight items over the analyzed period, the proposed research will focus on the UPC items but include the non-UPC items in the estimation of total purchases/sales.

Because of the data redesign, I separate the four-year period into two i.e. 2005-06 and 2007-08.[2] In addition, the separation of the two also allow us to analyze how consumer behavior responds to the Wal-Mart and other retailers' market expansion in organic market around the end of 2006. Within each set of two years, I use the first 26 weeks as the "initialization" period to identify shopper types and format-specific indexes to avoid potential endogeneity between organic consumption decision and store format choices. The remaining 78 weeks were used as the "estimation" sample. The estimation is based on every shopping trip of households with shopping duration being no longer than 30 days during the estimation period at the selected retail chains in the market, to ensure that each panelist was faithful in recording purchases and remained in the panel for the entire period.

The resulting data set had 710 households with a total of 45,877 shopping trips in 2005-06 sample and 942 households with 48,469 trips in 2007-08 sample. The selected retail chains for my analysis include 2 value-oriented retail chains consisting of 29 (37) stores, 4 traditional supermarket chains featuring 172 (147) stores, and 1 high-end specialty supermarket chain with 6 (7) stores in my 2005-06 (2007-08) sample.

Table 2 lists the definitions and descriptive statistics (means) of the explanatory variables used in the estimation of store format choice models for the markets of a non-coastal U.S. city in 2005-2008. One of the notable differences between the two samples is that I observe a fast growing pattern of organic consumption. In particular, the percentage of organic food purchase in total spending of the trip increases from 1.24% to 1.93%, which is with over 50% of growth. Furthermore, the percentage of frequent organic shoppers also increases from 3.33% to 4.68%, which is a 40% increase. We also learn that the average transaction amount recorded for each shopping

Table 2. Descriptive Statistics (Mean) of Variables for Store Format Choice Estimation, 2005-2008

Variable	Definition	2005-06	2007-08
price L/M	price index (value-oriented) / price index (supermarkets)	0.9754	1.0597
price H/M	price index (high-end) / price index (supermarkets)	1.5390	1.5438
total spending	total transaction amount recorded for the shopping trip	22.8651	18.1503
% dry grocery	% of dry grocery purchase in total spending of the trip	0.5284	0.3508
% fresh produce	% of fresh produce purchase in total spending of the trip	0.0657	0.0741
% organics	% of organic foods purchase in total spending of the trip	0.0124	0.0193
income	household income (in $1,000)	6.3300	6.8776
organic	1 if householder is a frequent organic shopper \ 0 otherwise	0.0333	0.0468
% discount	discount use rate during the initialization period	0.3421	0.3033
loyalty L	% of trips that household made to the format	0.1806	0.2245
loyalty M	(value-oriented, supermarkets, high-end)	0.8050	0.7636
loyalty H	during the initialization period	0.0144	0.0119
days between trips	number of days between two shopping trips	4.9563	5.4318
distance	the distance between consumer’s home and store	8.9900	9.3263
n(stores) L	number of value-oriented stores within 25 miles from household	1.1312	1.6564
n(stores) M	number of supermarket stores within 25 miles from household	5.7222	5.6494
n(stores) H	number of high-end stores within 25 miles from household	0.2619	0.2062
household size	number of persons in the household	2.3583	2.3924
less educated	1 if householder is not college educated \ 0 otherwise	0.1115	0.1081
single	1 if single householder \ 0 otherwise	0.4248	0.4275
preschool children	1 if family has child(-ren) under age of 6 \ 0 otherwise	0.0611	0.1127
school-age children	1 if family has child(-ren) aged 6~18 \ 0 otherwise	0.2290	0.2354
elderly	1 if householder is aged 65 and above \ 0 otherwise	0.2403	0.2320

Data source: A. C. Nielsen Homescan Data, and computed by this study.

trip decreased by about 20% over the two periods, likely reflecting the economy downturn occurred in the later period.

As to market shares, the data show that value-oriented format was gaining market shares, as its customer loyalty increases 24%. The market expansion of value-oriented format is also reflected in more number of stores over the two sample period. On the other hand, high-end specialty store format was losing its shares over time.

V. Results and Discussion

In this study, the store format choice consists of the choice of three alternatives, i.e. low-end value-oriented (L), middle format: supermarkets (M), and high-end specialty (H) formats. The alternative, the *middle format: supermarkets*, is chosen as the base outcome for estimation in the multinomial Logit Model setting. The analyzed samples are weighted by sampling weights, and heteroskedasticity-robust standard errors are calculated and reported. To account for differential price sensitivities among households, I first allow for price coefficients to vary among households. In addition, I interact the ratios of price indexes with household income and the income squared to capture the variation in price sensitivity due to the differences in WTP or quality perception associated with household income.

5.1 Parameter Estimates

Table 3 reports the MLE parameter estimates for store format choice from a mixed multinomial Logit Model. These parameter estimates represent the marginal utilities of the determinants, or equivalently the impacts of the determinants on log odds: $\ln\left(P_{hf}/P_{hM}\right)$, where f = L (value-oriented) or H (high-end specialty) is the store format choice in selection with format M (supermarkets) as the base.

Several key findings emerge from the regressions. First, the price sensitivity varies by the income level. The results from the parameter estimates of interaction term: *income*price* suggest that price sensitivity is greater for lower-income households, especially for those own-price terms. In addition, the parameter estimates from the price variables show statistically significant standard deviation, indicating price sensitivity varies among households. These results provide supporting evidence for hypothesis H1.

Table 3. MLE Parameter Estimates of Mixed Multinomial Logit Model

Variable		2005-2006				2007-2008			
		Value-oriented		High-end		Value-oriented		High-end	
		Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
		Random Parameters							
price L/M	mean	-4.238**	1.479	-7.779*	3.841	-1.396*	0.669	-0.142	2.615
	standard deviation	0.863**	0.213	0.805	0.508	0.759**	0.112	-0.177	0.384
price H/M	mean	-1.136	0.802	0.189	1.885	0.057	0.426	-0.417	2.329
	standard deviation	0.949**	0.242	0.456**	0.111	0.589**	0.060	0.601	0.362
ln(org L/M)	mean	-0.123	0.093	1.372**	0.494	-0.006	0.137	0.199	0.955
	standard deviation	0.148	0.120	1.718**	0.378	0.800**	0.193	1.348	0.864
ln(org H/M)	mean	0.320*	0.165	0.456	0.427	-0.323	0.210	-2.124*	0.983
	standard deviation	0.215**	0.066	0.062	0.086	0.190**	0.043	0.799**	0.267
income*price L/M		1.070*	0.445	0.770	1.174	0.073	0.141	-0.347	0.610
$income^2$*price L/M		-0.017	0.024	-0.066	0.074	0.005	0.007	-0.005	0.027
income*price H/M		0.136	0.240	-0.300	0.581	0.047	0.092	0.120	0.441
$income^2$*price H/M		0.002	0.015	0.040	0.043	-0.005	0.004	-0.006	0.019
income		-1.140**	0.243	-0.234	0.640	-0.157*	0.075	0.507	0.339
% discount		-3.427**	0.272	-5.786**	0.828	-3.204**	0.223	-5.665**	1.186
days between trips		0.006	0.009	-0.024	0.047	0.002	0.007	-0.073	0.054
distance		-0.032**	0.010	-0.010	0.009	-0.028**	0.005	-0.024	0.024
n(store) L		1.318**	0.191	0.371	0.312	0.740**	0.131	0.475	0.308
n(store) M		-0.127**	0.045	-0.157*	0.075	-0.213**	0.041	-0.348**	0.097
n(store) H		1.514**	0.313	3.137**	0.844	0.665**	0.219	5.135**	0.980
total spending		0.001	0.003	-0.001	0.008	0.002	0.003	-0.030	0.025
% dry grocery		1.924**	0.164	-0.520	0.597	0.604**	0.153	0.973	0.632
% fresh		0.367	0.377	0.827	0.909	0.263	0.252	2.058**	0.769
% organics		-2.820	2.019	6.448**	1.086	-1.311	0.775	4.258*	1.862
loyalty L		6.291**	0.460	-0.554	1.081	5.736**	0.291	2.439*	1.238
loyalty H		-1.370	1.448	15.379**	1.789	0.318	2.011	9.166*	3.625
organic shopper		0.099	0.580	0.701	0.479	0.086	0.310	2.056**	0.669
household size		-0.100	0.118	-0.117	0.499	-0.150	0.081	-1.857	1.007
less educated		0.057	0.385	-0.076	0.769	0.319	0.211	-24.693**	1.298
single		0.089	0.231	-0.320	0.942	-0.114	0.195	-1.318	1.142
preschool children		0.043	0.371	0.722	1.147	-0.179	0.246	2.370	1.341
school-age children		0.539	0.335	0.702	1.209	0.001	0.257	3.264	1.982
elderly		-0.506*	0.229	-0.241	0.410	-0.107	0.184	1.355	1.159

Data source: This study.

Note 1: Supermarkets (middle format) is the base outcome for all variables.

Note 2: Heteroskedasticity robust errors are reported.

Note 3: *, ** denote statistical significance at 5% and 1%, respectively.

Second, the estimated standard deviations for price coefficients are significantly different from zero for both equations and both sample periods. The sizes of standard deviations of price coefficients are uniformly larger in value-oriented equations than in high-end equations. It implies that variation in price sensitivity for the general formats' shoppers (value-oriented and supermarkets) is greater than the variation for the high-end format's shoppers. It supports the expectation in H2.

Third, the results from the *income* coefficients suggest that high-income families are more likely to shop at supermarkets than value-oriented stores, and at high-end than supermarkets. However, the latter is not statistically significant. In addition, the coefficients of *organic* shopper type in both periods confirm the prediction that frequent organic shoppers prefer high-end format retailers.

Fourth, the coefficients of quality indexes (log of the ratio of organic supply) represent the store image effects associated with consumer fixed utility component. The estimation result shows an interesting pattern of cross-format effects in the estimation of 2005-06 samples: an increase in quality ratio of value-oriented to supermarkets has positive effects on log odds of high-end formats versus supermarkets, and an increase in quality ratio of high-end to supermarkets also has positive effects on log odds of value-oriented versus supermarkets. In 2007-08, the own-format effect is significant in the high-end equation. Along with the observation that the difference between a value-oriented format's organic supply and the other two formats' was reduced over the two periods, the result suggests that own-format effect of a quality improvement arises only when quality difference between formats is small enough. In other words, less effective impact of quality may hinder quality improvement for a market with well distanced quality settings among formats. In addition, I observe statistically significant standard deviations in these random coefficients from both periods of estimation. This implies that the store image effect varies among households in the estimated samples.

Regarding to location and factors affecting shopping costs in store format choice, the results confirm the findings in the literature. In particular, the households with shopping patterns of more days between trips or longer distance tend to choose supermarkets over the value-oriented retailers or high-end specialty stores. This suggests that the households with higher shopping costs prefer formats featuring one-stop shopping environment-supermarkets, which feature broader assortments and more varieties in food categories.

The results from the coefficient of number of stores in each format within the shopping range of households suggest a network effect on consumer's shopping decision. As its store number increases, the format itself gains in the likelihood of consumer store visits. The cross-format coefficients indicate a complimentary relationship between value-oriented and high-end formats, as an increase in the number of stores of either format would have positive impacts on the probability for consumers to visit the other format. This implies that the low-high bundle of value-oriented and high-end formats appeals to households rather than other combinations, such as supermarkets and high-end formats together. From the estimates of loyalty L and loyalty H, we learn that the loyalty built from the past shopping experience is an important factor in explaining consumer's store format choice. The positive coefficient estimates of Loyalty L in 2007-08 High-end equation again support the argument that value-oriented and high-end market segments may be overlapping.

Furthermore, a household's store format choice decision is dependent of her basket size and composition. The estimation results show no significant difference in the impact of total spending per trip on the choice among the alternatives. Confirming the common belief, I find that households with higher expense share in dry grocery are most likely to shop at value-oriented stores than supermarkets, while organic food consumption helps the decision to shop at the high-end stores over the other two. This is also consistent with the implications from coefficient estimates of organic dummy, i.e. organic shoppers are more likely to shop at high-end stores than supermarkets, while they prefer supermarkets to the value-oriented format. The results do not show a differential shopping pattern for households of different types, from the parameter estimates of the demographic variables.

Finally, the scale of price sensitivity reduced significantly in value-oriented equation for 2007-08 compared to 2005-06, while price sensitivity in high-end equation increases over time. This is likely a result of increasing supply in organics due to Wal-Mart’s market expansion in the summer of 2006. The connection suggests that retailers’ increased supply on organic products have a positive impact on consumers’ quality perception and retailers’ store image.

5.2 Marginal Effect of Price

A direct interpretation of the parameter estimates reported in Table 3 provides only limited substantive behavioral meaning about the effect of specific factor on consumer’s store format choice outcomes. This arises when dependent variables appear as standalone variables and as nonlinear transformed variable and when they are interacted with other dependent variables. In these instances, a standard approach is to compute the full marginal effect of a dependent variable’s impact on the model, and the statistical significance of a utility parameter does not imply the same significance for the marginal effect (Greene, 2004). Therefore, I further compute the marginal effect of price for various income groups of households.

A marginal effect, defined as the derivative of the probability, is the influence a one unit change in an explanatory variable has on the probability of selecting a particular outcome. Under the present model setup, the effect of a change in price involves three terms: 1) main term (price), 2) income interaction term (income*price), and 3) income squared interaction term ($income^2$*price). Therefore, the marginal effect of price would vary among income groups as a result of these effects combined. Furthermore, the price effect will vary among households, regardless of their income, since the Mixed Logit Model allows random parameters for the main term of price.

Let y denote the dependent variable, which is a dummy variable of store format choice. The conditional mean of the dependent variable is

$$E\left[y \mid p, m, X\right] = L\left(\beta_1 p + \beta_2 mp + \beta_3 m^2 p + X\beta\right) \equiv L\left(u\right) \tag{7}$$

where $L(\cdot)$ is the logistic function and u denotes the index $\beta_1 p + \beta_2 mp + \beta_3 m^2 p + X\beta$, p denotes price and m denotes income. Suppose that price and income (p, m) are both continuous. The marginal effect of price is thus

$$\frac{\partial L\left(u\right)}{\partial p} = \left(\beta_1 + \beta_2 m + \beta_3 m^2\right) L'\left(u\right) \tag{8}$$

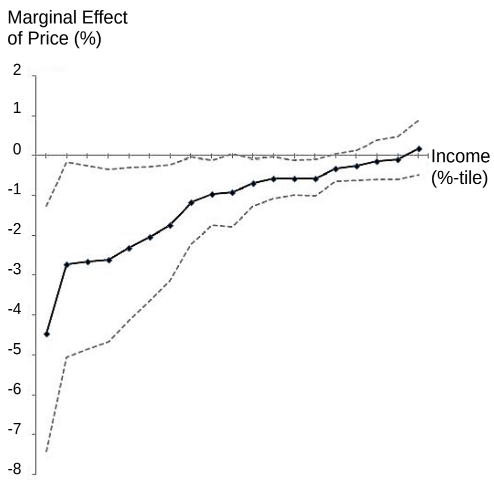
Unlike linear models, the marginal effect consists of $\beta_1 + \beta_2 m + \beta_3 m^2$ and $L'\left(u\right)$.

Figure 2 presents the marginal effects for various income groups, by plotting the percentage change in predicted probabilities for a percentage change in prices for value-oriented and high-end equations for both periods of samples, where the horizontal axis is income in percentile and the vertical axis is the marginal effect of price.[3] The solid lines represent the means of marginal effects of price calculated from individual household's marginal effect, and the dotted lines represent the upper and lower bounds of confidence intervals (95% & 5%).[4]

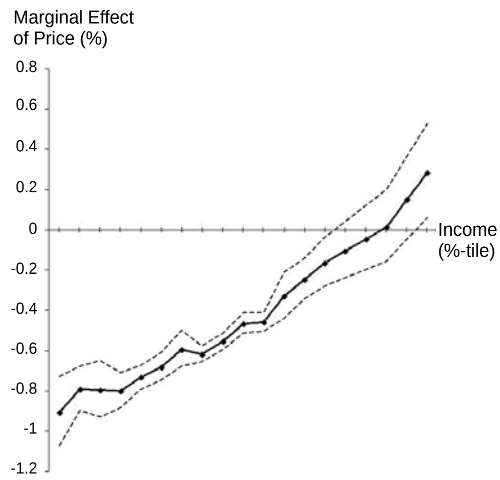
As shown in the figures, those results from comparison of marginal effects of price among households and across formats support the hypotheses:

H1. Price sensitivity varies among households of different types and among formats, and

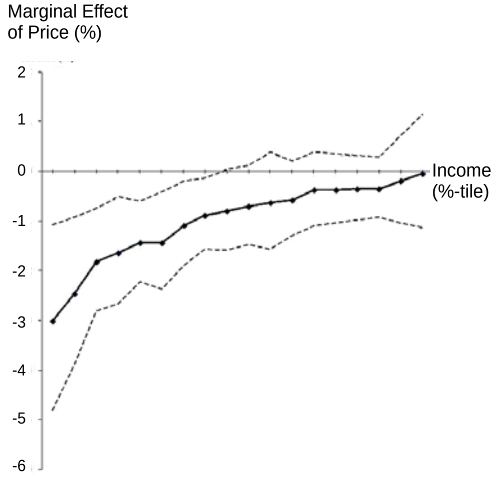
H2. Variation in price sensitivity for the general formats' shoppers is expected to be greater than the variation for the high-end format's shoppers.



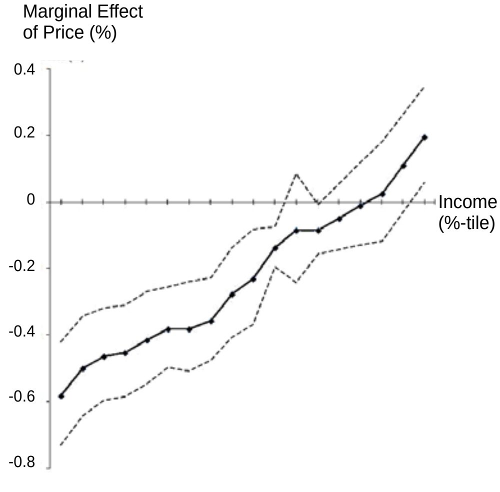
(a) Value-oriented, 2005-06



(b) High-end, 2005-06



(c) Value-oriented, 2007-08



(d) High-end, 2007-08

Figure 2.　Price Effect on Store Format Choice Probability by Income Groups

Data source: This study.

First, the marginal effects of price vary among income groups. Particularly, the magnitude of marginal effect decreases as income increases. For those rich households, the marginal effect of price is even greater than zero at the mean values and even the entire confidence interval for the top income groups. For value-oriented equation (versus supermarkets) in the sample of 2005-06, the marginal effects vary from -4.475 to 0.176, and I observe significant differences in standard deviation across income groups: the variation of marginal effect is much greater for the poor than for the rich. I observe an upward-sloped curve for high-end equation in 2005-06 as well, but the variation in marginal effect is maximized at the top income groups instead. Upward-sloped curves are observed for both value-oriented and high-end equations in 2007-08 samples too, although the variations in marginal effects are more uniformly sized across income groups. In sum, those results support hypothesis H1.

The marginal effects of price differ across formats as well: the effects in value-oriented equation are about 2-3 times of those in high-end equations. This provides support for the latter part of hypothesis H1. Next, the variation in marginal effect for high-end shoppers is relatively smaller than that for value-oriented shoppers as well. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, is 0.27 for high-end and 0.90 for value-oriented in 2005-06; 0.57 for high-end and 0.92 for value-oriented in 2007-08. It thus provides supporting evidence for hypothesis H2.

VI. Conclusions

In this chapter, I use a unique household purchase panel data to investigate the determinants of consumer store format choice decision making and examine the theoretical predictions from chapter 2. Several key findings emerge from the analysis of a non-coastal U.S. metropolitan area in 2005-06 and 2007-08. First, I find that household's price sensitivity in store format choice varies among households and decreases with income. The variation in price sensitivity is relatively noteworthy for the general formats' shoppers compared to high-end shoppers. Second, organic food demand and supply are both important factors influencing households in choosing the high-end specialty format. In particular, the estimated random parameters for organic supply ratio show statistically significant variation in store image effect among households and the effect due to own format's quality improvement is significant only when quality settings are more comparable among formats. On the demand side, the results suggest that organic food consumption is an important reason leading consumers to choose a high-end format over the other two formats. Third, the results from the network effect of the number of stores in the area suggest a complementary relationship between value-oriented and high-end formats, a likely pair of inferior and superior good combination, while supermarkets format presents the features of an one-stop shopping outlet. In addition, I find supportive evidence that shopping costs measured by days between trips and the distance between store and household are important to consumers' format choice between the specialty format and one-stop shopping formats, confirming the common findings in the literature. Finally, basket size and composition on the shopping list, format loyalty and shopper types are also found to be important to a household's store format choice decision making.

In sum, my findings verify key predictions from the theory. First, the shoppers of general formats (value-oriented stores and supermarkets) have more diversified preferences than those that shop at high-end specialty formats. Moreover, high-end specialty format shoppers have higher WTP and quality perception compared to general format shoppers. Second, the *value* factor, defined as the ratio of quality to price, is central in the theoretical model in determining the consumer store format choice decision. Since value is the deterministic factor, both price and quality effects are considered empirically. Even with a quality proxy like the organic supply rate used in this study, the estimated price effect may still contain partially the impact due to unobserved difference in quality perception among households. As a result, observed price sensitivity should vary among consumers and among formats. And indeed, the estimation results verify the prediction.

Endnotes

1. For the choice with multiple alternatives (greater than two) like the present one, the approach works equally well after creating duplicate records of data (see Hole (2004), pp. 397-398).
2. Organic information is not available for random-weight fresh produce for 2007-08 data set. This will likely lead to underestimate consumers' organic penetration, as fresh produce has always been on the top of consumers' shopping list for organics.
3. In estimation, one percentage change in price was applied to compute the difference in predicted probabilities between with and without the change in price for each individual. The difference in predicted probabilities was then divided by the predicted probability before change to be reported as the marginal effect due to a 1 % change in price. This measure is actually in a form of elasticity.
4. The upper bound is valued at (mean + 1.05*sd) and the lower bound is valued at (mean - 0.96*sd) for our sample, of which size is over 1000.

References

Baker, J., D. Grewal, and A. Parasuraman, 1994. "The Influence of Store Environment on Quality Inferences and Store Image," *Journal of the Academy of Marketing Science*. 22 (4): 328-339.

Basker, E. and M. Noel, 2007. "The Evolving Food Chain: Competitive Effects of Wal-Marts Entry into the Supermarket Industry," Working Paper No. 2007-03. Department of Economics, University of California, San Diego.

Basker, E., 2005. "Selling a Cheaper Mousetrap: Wal-Mart's Effect on Retail Prices," *Journal of Urban Economics*. 58: 203-229.

Bell, D. R. and J. M. Lattin, 1998. "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shopping Prefer EDLP," *Marketing Science.* 17: 66-88.

Bell, D. R., T. Ho, and C. S. Tang, 1998. "Determining Where to Shop: Fixed and Variable Costs of Shopping," *Journal of Marketing Research*. 35: 352-369.

Bhatnagar, A. and B. T. Ratchford, 2004. “A Model of Retail Format Competition for Non-durable Goods,” *International Journal of Research in Marketing.* 21: 39-59.

Briesch, R. A., P. K. Chintagunta, and R. L. Matzkin, 2010. “Nonparametric Discrete Choice Models with Unobserved Heterogeneity,” *Journal of Business and Economic Statistics.* 28(2): 291-307.

Brunsø, K., T. A. Fjord, and K. G. Grunert, 2002. “Consumer’s Food Choice and Quality Perception,” MAPP Working Paper No. 77. The Aarhus School of Business, Aarhus, Denmark.

Carpenter, J. M. and M. Moore, 2006. “Consumer Demographics, Store Attributes and Retail Format Choice in the US Grocery Market,” *International Journal of Retail and Distribution Management.* 34: 434-452.

Dodds, W. B., K. B. Monroe, and D. Grewal, 1991. “Effects of Price, Brand, and Store Information on Buyers’ Product Evaluations,” *Journal of Marketing Research*. 28: 307-319.

Fox, E. J., A. L. Montgomery, and L. M. Lodish, 2004. “Consumer Shopping and Spending across Retail Formats,” *Journal of Business*. 77: 25-60.

Franklin, A. W., 2001. “The Impact of Wal-Mart Supercenters on Supermarket Concentration in U.S. Metropolitan Areas,” *Agribusiness*. 17: 105-114.

Gotlieb, J. B., D. Grewal, and S. W. Brown, 1994. “Consumer Satisfaction and Perceived Quality: Complementary or Divergent Constructs? ” *Journal of Applied Psychology*. 79 (6): 875-885.

Greene, W., 2004. “Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models,” Working Papers 04-08, Department of Economics, Stern School of Business, New York University, New York, NY.

Ho, T., C. S. Tang, and D. R. Bell, 1998. “Rational Shopping Behavior and the Option Value of Variable Pricing,” *Management Science*. 44: 145-160.

Hole, A. R., 2004. “Fitting Mixed Logit Models by Using Maximum Simulated Likelihood,” *The Stata Journal.* 7: 388-401.

Hsieh, M. F., 2012. “Store Format Choice,” PhD Dissertation, the University of Wisconsin-Madison, Department of Agricultural and Applied Economics, Madison, WI.

Parasuraman, A. and D. Grewal, 2000. “The Impact of Technology on the Quality-Value-Loyalty Chain: A Research Agenda,” *Journal of the Academy of Marketing Science*. 28 (1): 168-174.

Stiegert, K. W. and T. Sharkey, 2007. “Food Pricing, Competition and the Emerging Supercenter Format,” *Agribusiness*. 23: 295-312.

Train, K., 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Voss, G., A. Parasuraman, and D. Grewal, 1998. “The Role of Price and Quality Perceptions in Prepurchase and Postpurchase Evaluation of Services,” *Journal of Marketing*. 62: 46-61.

異質商家型態消費選擇之價格效果*

謝銘逢**

價格是消費者之購物決策考慮之要素，因此了解及估算消費選擇之價格效果為市場分析及經濟研究之重要課題。本實證研究利用 A.C. Nielsen 的家戶消費資料，分析美國消費者於 2005-2008 年間對食品雜貨之異質商家型態選擇，並探討價格如何影響消費者之決定。混合多項邏輯模式之迴歸分析結果顯示，相對於一般的消費族群而言，高價特質商家之消費族群對價格認知較無差異。此外，結果亦顯示，不同所得水平的消費群對價格的敏感度不同，相對低收入家戶而言，高所得家戶對價格較不敏感。本實證方法與結果有助於異質偏好消費者對異質商家型態消費選擇之價格效果的了解。

關鍵詞：價格效果、商家型態選擇、混合多項邏輯模式、超級市場、食品零售業

* 本人在此感謝 Kyle W. Stiegert、Jean-Paul Chavas、Timothy J. Richards、Pual D. Mitchell 和 Brian W. Gould 等教授對本文提供了建設性的意見與建議。我也很感謝美國農業部經濟研究局的 Ephraim Leibtag 博士和 Biing-Hwang Lin 博士，及 A. C. Nielsen 公司提供本研究之關鍵資料。感謝兩位匿名審查人對本文提供之幫助與建議，在此特別致謝。

** 國立臺灣大學生物產業傳播暨發展學系助理教授。