10.6196/TAER.202006_26(1).0001

Dietary Choices and Income Inequality: Evidence from Household Food Purchase Data*

Ming-Feng Hsieh**

Inequality has become one of the most renowned features of capitalism. This paper focuses on the association between income inequality and people's dietary choices by documenting the differences in dietary patterns between metro markets and their connection to food consumption and income distributions. This empirical study uses a U.S. multi-market household food purchase data set to examine the influence of socioeconomic and demographic factors on health-related dietary choices of salt-sugar-fat intake, organic, vegetable, alcohol and tobacco use. Regression results show that health awareness, ability to pay, price, shopping pattern, and market environment are key to affect individual dietary choice. The effects of income on diet are found not only at the individual level, but also through aggregate income per capita and the degree of inequality. This study provides useful insights for understanding the rationales behind complex dietary behavioral decisions and their distributional differences.

* This project has benefited from financial support by the Ministry of Science and Technology, Taiwan (103-2410-H-002-026/106-2410-H-029-010). The author thanks the editor and two anonymous referees for helpful comments and suggestions.

** Assistant Professor, Department of Economics, Tunghai University.

Received 30 March 2020; Received in first revised form 1 May 2020; Received in second revised form 7 May 2020; Accepted 21 May 2020.

農業經濟叢刊 (Taiwanese Agricultural Economic Review)，26:1(2020)，1-28。
臺灣農村經濟學會出版

Inequalities and dietary choice have important welfare meanings to all societies, not just the United States but also Taiwan or other countries. This study uses a large-scale and comprehensive transactional data which makes it possible to demonstrate how to decompose household food purchase records for further understanding of the association between dietary choice and income inequality. With the approach and insights learned from the case of the U.S., policy makers can utilize the information of income distribution and its potential associations with individual health-related dietary choices to design relevant public health policy or nutritional subsidy program.

***Keywords**: dietary choice, income distribution, inequality, food consumption, food purchase decision*

***JEL Codes:** I14*

I. Introduction

Dietary choice and selection of foods are the decisions that individuals make every day. It may be insignificant in terms of any single decision; however, the implications of dietary choice to an individual or the society are very influential. It is well known that the choice of foods has critical impacts on an individual's health and general wellbeing. The dietary style also reflects or are affected by the individual's income, education and general lifestyle. To the society, the distribution of individual diet may have distinct policy implications that differ from the distribution of other demographic variables such as income. The multi-dimensional nature of dietary patterns adds great complexity to the task of deciphering the behavior and its consequences. For instance, consuming greater quantity of foods does not necessarily coincide with greater quality, and a substitution between the two may be hidden behind a constant household expenditure in foods. Furthermore, high quality of food, often referred by its market price or value, is not always translated to a healthier diet. A society of the wealthiest individuals is not necessarily the healthiest in terms of diet. The complexity motivates this research to document and learn from what the distribution of diet and food consumption represents, its association with income distribution, and the policy implications.

This research is related to the literature in the following four aspects: (1) inequalities in consumption and income, (2) food choice and consumer behaviors, (3) diet and health, and (4) cross-sectional difference and development. First of all, income and consumption are closely connected in either individual or aggregate level. For individuals, consumption is a key motive of working and earning, while the ability to consume is constrained by the budget, i.e. income, on the other hand. The issue of increased income inequality, i.e. income is distributed more unevenly in the society, continues to influence policies, markets and household behaviors. Within the literature on inequality, most studies suggest that consumption is much equally distributed

compared to income (Blundell, Pistaferri, & Preston, 2008; Blundell & Preston, 1998; Hassett & Mathur, 2012; Meyer & Sullivan, 2009, 2010; Krueger & Perri, 2006), while some others find similar increasing trends of inequalities in consumption and income (Attanasio, Battistin, & Leicester, 2006; Attanasio, Hurst, & Pistaferri, 2012; Fisher, Johnson, & Smeeding, 2013). While the findings provide us insights about inequality matter, food consumption and diet were rarely at the center stage of study in literature.

The determinants of food choice and the related consumer behaviors have driven a vast majority of research interests in studying consumer demand, food system, store choice, market structure, marketing strategy from economic, marketing, sociological and psychological perspectives (e.g. Baumgärtner, Drupp, Meya, Munz, & Quaas, 2012; Bell, Ho, & Tang, 1998; Berry & Haile, 2009; Furst, Connors, Bisogni, Sobal, & Falk, 1996; Hsieh, 2012; Hsieh, Mitchell, & Stiegert, 2009; Hsieh & Stiegert, 2012; Kim, Allenby, & Rossi, 2002; Reutterer & Teller, 2009; Song & Chintagunta, 2007; Tang, Bell, & Ho, 2001). These studies are more interested in understanding how consumers make their purchase decisions, rather than in the making of the distribution and its related impacts.

Another aspect of research looks into the linkages between food consumption and health. For example, Boumtje, Huang, Lee and Lin (2005), Chang and Nayga (2009), Dougkas, Reynolds, Givens, Elwood and Minihane (2011), Lin, Huang and French (2004), Van Strien, Herman and Verheijden (2009) and Ver Ploeg and Ralston (2008) studied how dietary habits and other food-related factors affect body weight and the likelihood of obesity. Some other studies have also shown that communication, consumer knowledge on nutritional information, and nutritional quality have great influence on consumer's food choice for both at home and away from home consumption (e.g. Golan, Kuchler, Mitchell, Greene, & Jessup, 2001; Kreuter & Wray, 2003; Lin & Gruthrie, 2012; Saarela, 2014; Verbeke, 2008; Yen, Lin, & Davis, 2008). Another cluster of researchers, e.g. Allison and Foster (2004), Madden (2010), Naga and Yalcin (2008, 2010), and Van Doorslaer and Jones (2003), have looked into income-related inequality in health mainly using health survey data like National

Population Health Survey. Their measurement and quantitative analyses are limited to income and health issues but no connection with food consumption.

Finally, individuals' diet and their health outcomes progress differently across locations, over time, and across demographic groups. It is shown that food demand, nutritional intakes and health outcomes differ among countries at different stages of economic development (Pitt, Rosenzweig, & Hassan, 1990; Strauss & Thomas, 1998; Winters, Essam, Zezza, Davis, & Carletto, 2010). Food purchases differ substantially among regions within or among countries. Dubois, Griffith and Nevo (2014) documented the differences in food purchases and nutritional outcomes across three developed countries, the United States, United Kingdom, and France. It is expected to see differences in diet or food consumption in different years of time too, provided that technology, life style, eating habit or social value has changed compared to last decade or even just a year ago.

This study focuses on the association between income inequality and individual's dietary choices. The empirical study utilizes a unique household panel data of food purchase from 23 metro markets in the United States for a three-year period from 2006 to 2008 to demonstrate how to gain further understanding of dietary choice and income inequality through people's food transaction records. The rich information contained in this household food purchase dataset makes possible to identify and discuss the distributional differences in dietary styles, food consumption behaviors, and income among various markets. In addition to exploring the distributional measures of diet and income, the study further identifies the possible determinants responsible for the differences in diet and food consumption. Especially, the effects arisen through aggregate measures of income, average income per capita and the degree of income inequality, on household's food consumption and dietary behaviors are analyzed. These results from the case of the U.S. shall help us understand and address the dietary choice and inequality issues in Taiwan or other countries.

The remainder of this paper is organized as below. Section II documents the

distributional patterns of income, food consumption and diet, with an emphasis on the distributions of healthy versus unhealthy diet or lifestyles, and their associations with income. Section III discusses the results from regression estimation with both individual and aggregate measures of income to examine the determinants of household dietary styles. The paper concludes with a summary of findings and remarks.

II. Diet, Food Consumption and Income Distribution

The primary dataset used for this study is a panel dataset: the Homescan data (Nielsen, 2008) in major metropolitan markets that covers a three-year period between 2006 and 2008. The data panel includes a representative of households in 52 (metro) market areas and nine remaining (rural) areas in the 48 continental states of the US. This study uses a subset of data from the 23 larger metropolitan markets to have sufficient numbers of observations for a better presentation of distributional patterns. The data were collected through a household scanner process, in which panelists record their food and dry grocery 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 used for the analysis are from those in a so-called "static" panel, including the households who had participated in at least 10 of the 12 months during the year they participated. In terms of recording rate, there were over 95% of panelists having reported at least one shopping record weekly according to the calculation. This is a useful dataset to document the distribution of food consumption as it contains adequate information of household purchases from all sorts of grocery stores and retail outlets. Since the data are recorded from the households' purchase transactions, they are relatively more accurate and representative for households' actual choice of foods and dietary patterns than those based on the so-called "recall data" from a conduction of survey.

For baseline understanding of income and consumption inequality, several waves of data from the Consumer Expenditure Survey (CE) are adopted. The CE dataset consists of two components: The Interview Survey and the Diary Survey. The Interview Survey collects the information for larger purchases in households, such as spending on health services, durables, rents or utilities, once per quarter for four consecutive quarters. The Diary Survey on the other hand asks each panelist to keep a diary for two one-week periods recording small, frequently purchased items, such as spending for food, beverages, tobacco, personal care products, and nonprescription drugs and supplies. Approximately 7,000 usable interviews are collected from the Interview Survey each quarter and approximately 14,000 usable diaries are collected for the Diary Survey per year, based on data from year 2010 (Bureau of Labor Statistics, 2016). It is designed to be representative of the entire U.S. population, but not for cross-sectional studies for areas smaller than states.

Table 1 documents the income and consumption patterns in the United States in recent decades. The table describes the income and selected consumption spending patterns, including the means, standard deviations and the Gini coefficients that measure statistical dispersions, for the U.S. samples in 1996, 2000, 2005 and 2010. Over the years, the means of income, total consumption and other spending maintain an increasing trend, except for transportation and apparel spending in recent years. Although it is observed a relatively stable movement in the degree of dispersions, measured by Gini coefficient, between the waves of samples, the distributional patterns of spending vary between subcategories. Spending in food away from home, apparel, transportation, education and entertainment are with a larger Gini (ranged from 0.53 to 0.77), while spending in housing, food at home, food total and total consumption are with a rather equitable distribution (Gini ranged between 0.40 and 0.48). In particular, the pattern of total food or food at home spending is more equitable than distribution of income.

Table 1. Income and spending patterns in the United States, 1996-2010

	1996		2000		2005		2010	
Variable	Mean (S.D.)	Gini	Mean (S.D.)	Gini	Mean (S.D.)	Gini	Mean (S.D.)	Gini
Income	28,289 (32,939)	0.48	32,437 (40,583)	0.48	56,601 (55,956)	0.46	61,408 (60,738)	0.46
Consumption total	4,909 (5,440)	0.47	5,912 (6,530)	0.48	7,073 (7,741)	0.47	7,776 (8,417)	0.46
Food total	755 (709)	0.42	870 (798)	0.42	932 (860)	0.42	1,173 (1,076)	0.42
Food at home	552 (457)	0.40	632 (541)	0.41	682 (573)	0.41	794 (669)	0.40
Food away from Home	204 (447)	0.58	238 (462)	0.56	250 (498)	0.58	379 (666)	0.53
Apparel	202 (433)	0.63	252 (544)	0.63	226 (578)	0.63	192 (576)	0.63
Housing	1,588 (1,648)	0.47	1,953 (2,230)	0.48	2,285 (2,578)	0.47	2,660 (3,045)	0.46
Transportation	993 (2,932)	0.72	1,200 (3,485)	0.72	1,315 (3,843)	0.71	1,214 (3,370)	0.67
Education	84 (557)	0.73	94 (650)	0.74	144 (995)	0.73	172 (1,531)	0.77
Entertainment	268 (1,706)	0.67	315 (797)	0.63	362 (1,238)	0.63	392 (1,494)	0.61
No. of Households	4,971		9,718		8,900		7,107	

Data Source: Bureau of Labor Statistics (2016) (author's calculations added)

The rich information on food purchases from the Homescan data makes it possible to further investigate health-related dietary behavior. From the product selection and purchase outcomes, the study identifies whether the household is pursuing a healthy or unhealthy diet or lifestyle. 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, data contain some nutrient facts for products, like low-salt or low-fat, which are employed in this study to gain understanding of dietary styles in household food at-home consumption. The low-salt

diet identifier is defined as the percentage of low-salt option being chosen for the food items that provide low salt and regular options. The low-fat diet measure is based on household's choice of dairy products, like milk and yogurt. It is also a percentage of low-fats were purchased for the items with low fat and whole options. The sugar measure highlights the percentage of consuming foods with high sugar contents, such as soft drinks, disserts.

In the study, the consumption of organics, vegetables and the foods with low salt, low sugar, or low fat contents are considered as healthy, while the consumption of alcohol and tobacco use are unhealthy. Two measures are employed: (1) the Gini measure to demonstrate the degree of dispersion for each variable, and (2) a modified Concentration Index (C.I.) proposed by Erreygers and Van Ourti (2011) to indicate the association of the distribution of specific variable with income. The specific C.I. formula, which captures the covariance between individual consumption (c_i) and individual relative income rank (R_i), is as below:

$$C.I. = \frac{2}{N^2\left(\mu_c - \min\{c_i\}\right)} * \sum_i c_i R_i \tag{1}$$

where c_i is the household i's consumption on the good of interest, R_i is the relative income rank of the household among the population of N households, μ_c and $\min\{c_i\}$ denote the mean and minimum of c in all households. The index measure is modified to be ranging between -1 and +1, reflecting negative or positive association between individual consumption and income rank.

Table 2 summarizes the purchase patterns of health-related diet based on a pooled sample of all markets. The C.I. numbers show that the consumption of low-salt, organics, vegetables and alcohol are positively associated with income, while low-fat and tobacco products are negatively associated with income. These results provide some insights for further investigation. For example, it suggests that the lower-income households have higher tendency to purchase low-fat options, such as milk. In practice,

nonfat or low-fat milk is cheaper than regular milk, thus low-fat option is more appealing to the lower-income. Hence, price effect is critical to the distributional outcome observed too. Sugar intakes are not associated with income in a clear direction. It is consistent with the general finding in the literature: craving for sugar does not link to any specific income group, although some research shows that lower-income individuals tend to consume more foods with high-sugar contents. The Gini coefficient estimates indicate that the consumption of low-salt, low-fat intakes, organics, alcohol and tobacco use has a greater dispersion, while vegetable and sugar intakes are distributed more equally among households.

Table 2. Health-related dietary patterns and income in the U.S., 2006-2008

Household Consumption	N(obs.)	Spending Share Mean (S.D.)	C.I. with Income Index (S.D.)	Gini coefficient
Food at home total	76,658		0.069 (0.001)	0.32826
low salt diet	75,243	0.017 (0.059)	0.028 (0.007)	0.53734
low fat diet	75,184	0.208 (0.296)	-0.084 (0.003)	0.56683
low sugar diet	76,658	0.677 (0.106)	-0.001 (0.001)	0.18149
Organic	76,658	0.010 (0.029)	0.153 (0.006)	0.71334
Vegetable	76,658	0.095 (0.053)	0.004 (0.001)	0.29549
Alcohol	76,658	0.036 (0.082)	0.056 (0.005)	0.67212
Tobacco use	76,658	0.008 (0.045)	-0.201 (0.012)	0.70288

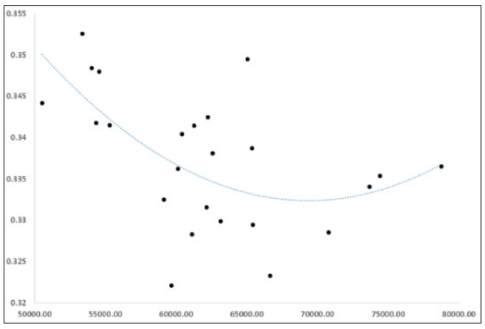
Data source: Nielsen (2008) (author's calculations added).

To further examine the patterns of consumption by market, the Gini measures of income, total food spending, and each individual consumption are plotted against the average income levels of markets. The cross-sectional comparisons among markets as in Figure 1 show that organics, alcohol and tobacco use are with greater inequality measures (Gini) than the income's, and the size of dispersion for this group of consumption goods is negatively correlated with market's income level. While total food-at-home consumption, vegetables are with smaller distributional dispersion, the inequality measures increase with market income.

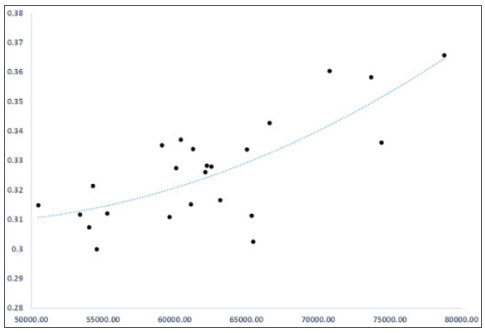
III. Household Dietary Choice

This study examines the determinants of households' dietary choices relevant to health matters, including whether their diet tends to have high percentage of foods with low-salt, low-sugar and low-fat contents, and their consumption of alcohol, tobacco products, organics, and vegetables. The measurements of health-related dietary choices examined in this study are as follows. First, a score measure for low salt-sugar-fat dietary choice (**low salt-sugar-fat diet**) is defined as

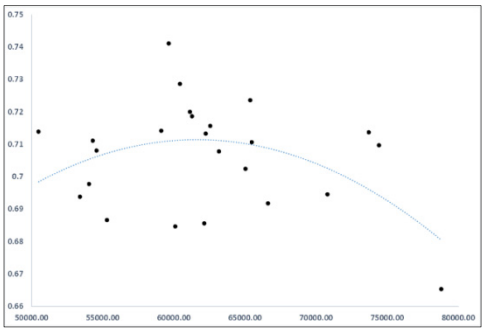
$$\text{score}_i = \sum_k \frac{r_{ki} - \overline{r}_k}{r_k}, \tag{2}$$



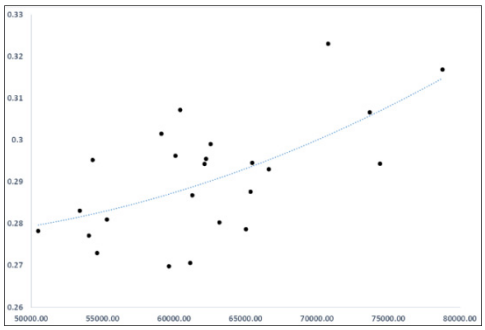
(1) Income



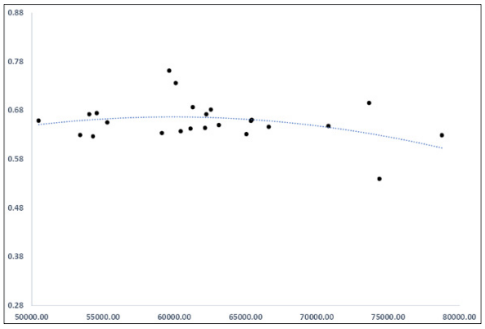
(2) Food at home total



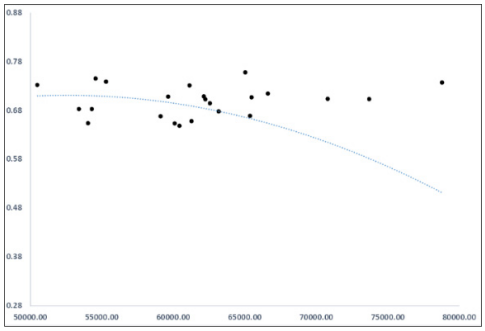
(3) Organic



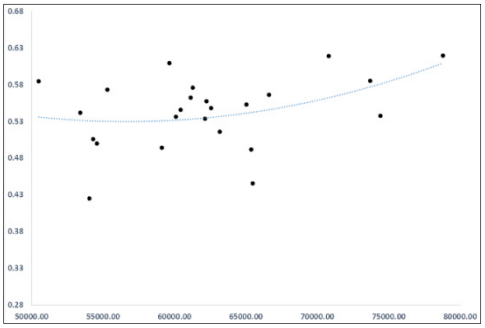
(4) Vegetables



(5) Alcohol



(6) Tobacco



(7) Low-salt-sugar-fat diet

Figure 1. Inequalities in diet, food consumption and income by market

Data source: Nielsen (2008) (author's calculations added).

Note: Vertical Axis represents the Gini value and horizontal axis is the average income. Each point represents one of the 23 selected markets.

where the subscript k indicates the type of low-nutrient, including low-salt, low-sugar, and low-fat; i represents individual household; r_{ki} is an individual household i's ratio of consuming low-nutrient option of foods and $\overline{r_k}$ is the aggregate mean of the associated ratio. The higher score indicates a bigger tendency for the household opting a low salt-sugar-fat dietary style.

The second diet indicator (**organic**) to be examined in the regression is the organic penetration rate, which is defined as the total spending in organic food products to the total food spending. Organic products are known as healthier and more

environmental-friendly food products to consumers. They are normally priced with a premium compared to their conventional counterparts. In addition, another three indicators of individual's health-related dietary choices: the percentages of food purchases in vegetables (**vegetables**), alcohol (**alcohol**), and tobacco products (**tobacco**) are measured and included as dependents to be examined in the regression.

Since the score measure for low salt-sugar-fat dietary choice is continuous and ranging from negative values to positive ones, the estimation is performed with ordinary least square regression and fixed effect approach. The other four equations of regression, with the dependents being spending shares in organics, vegetables, alcohol, or tobacco, which are with a certain ratio of zero-value observations. Specifically, organic equation has 27,659 (or 36.08%), vegetable equation has 273 (or 0.36%), alcohol equation has 33,725 (43.99%), and tobacco equation has 68,011 (or 88.72%) left-censored observations. Therefore, it is appropriate to employ the Tobit regression model (Tobin, 1958) with left-censored observations for estimation. In specific, the Tobit model is a censored normal regression model, of which the estimation is done with the maximum likelihood method[1]. The Tobit model is defined as follows:

$$y_i = \beta' x_i + u_i \quad \text{if RHS}>0, \tag{3}$$

$$y_i = 0 \quad \text{otherwise,}$$

where β is a vector of unknown parameters; x_i is a vector of known explanatory variables; u_i are residuals that are independently and normally distributed with mean zero and a common variance σ^2.

Two model specifications are estimated in the analysis. In the first specification - the fixed-effect (FE) model, the explanatory variables include the demographics, the shopper types, and household income, with controlling the market-specific fixed effects in estimation. In this setup, no market characteristic is modelled in and so there might be some unobserved time-invariant effects associated with markets. The FE model eliminates the unobservable by demeaning the variables using the within

transformation. This study uses the dummy-variable approach by adding a dummy variable for each market to control for the market-specific fixed effects. Except for the equation of low salt-sugar-fat dietary choice being standard FE model, the other four equations are estimated with the Tobit model setup with market fixed-effects controlled (Tobit-FE). In the second specification – the market-factor model, a set of market-level factors are added as explanatories, instead of simple fixed effects for markets. Specifically, market average income, market income inequality, market share of value-oriented and population, are added to the set of explanatory variables in addition to the demographics, shopper types and income of households. The estimation for the first equation (low salt-sugar-fat diet) is done with the least square regression, while the rest four equations are estimated with the Tobit regression. In all regression estimations, heteroskedasticity-robust (robust) standard errors are applied. As an effort of minimizing potential endogeneity issue, the study uses the first 52 weeks (of the three-year data period) as the initialization period for constructing the measures of shopper types. The remaining 104 weeks were used as the estimation sample[2]. The resulting sample for regression estimation consisting of 76,658 households observations over a period of three years between 2006 and 2008[3].

3.1 Explanatory Variables

Three sets of variables are employed to explain the dietary choice of households: (1) demographic variables, (2) shopper types, and (3) income factors. 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. internet accessibility). The specific household demographic characteristics employed in this study are: household size (number of members), education of the householder (1 if college educated or 0 else), marital status of the householder (1 if married or 0 else), with preschool age children (age less than 6), with school-age children (ages 6-18), and

elderly (1 with householder aged over 65).

Second, the study uses the following measures of shopping styles to reflect the shopper types of households, namely shopping frequency (number of shopping trips), average basket size (average spending per trip), discount use rate (frequency), and value-oriented shopper. Since these general shopping styles coincide with the dietary choices, the shopper types are generated with the data in the "initialization" period (the first 26 weeks) to avoid potential endogeneity. The shopping frequency is defined as the average number of shopping trips. The average basket size is measured by the average spending per trip. The household discount use rate is calculated by the ratio of items purchased with coupons or discounted prices to total number of items, to capture their sensitivity to price changes. The study uses a dummy variable with the value of 1 if household's favorite retail channel is "mass merchandise" or "supercenter" to categorize as a value-oriented shopper. As shown in the literature, for example Hsieh and Stiegert (2012), retail format choice is strongly linked to consumer's food choice.

Three income measures relevant to a household and the market it belongs to, namely (1) household income, (2) market average income, and (3) market income inequality, are employed to further understand how each of these income factors play a role in household food at home purchase and consumption. Individual household income is certainly a key factor to household dietary choice as it determines the ability to pay. Hence, it is included in both specifications of estimation. The average income level and the degree of income inequality in the market where the household resides can help us to gain further understanding of the potential impacts of underlying economy performance and distributional resource allocation on individual food choice. In addition, market share of value-oriented retailers and population are included in the market-factor models to reflect food retailing and locational market environment.

Table 3. Descriptive statistics for the households, 2006-2008

	Mean (S.D.)		Mean (S.D.)
Number of households	76,658	Household size	2.343 (1.262)
Score (low salt-sugar-fat diet)	0.001 (3.711)	Some college educated	81.1% (0.392)
% low-salt diet	1.7% (0.059)	Married	60.0% (0.490)
% low-sugar diet	67.7% (0.106)	Preschool children	6.6% (0.248)
% low-fat diet	20.8% (0.296)	School-age children	18.9% (0.392)
Organic penetration rate	1.0% (0.029)	Elderly	29.2% (0.455)
Spending % on vegetables	9.5% (0.053)	Shopping frequency	1.66 (0.899)
Spending % on alcohol	3.6% (0.082)	Average basket size	19.15 (12.901)
Spending % on tobacco	0.8% (0.045)	Discount use rate	33.8% (0.268)
Income ($10,000s)	6.224 (4.041)	Value-oriented shopper	10.2% (0.303)

Data source: Nielsen (2008) (author's calculations added).

3.2 Descriptive Statistics

Descriptive statistics of the analysis sample are provided in Table 3. The index of score for quantifying low salt-sugar-fat diet is averaged around 0 with a standard deviation of 3.7108, ranging from -3.5843 to 60.9467. In details, the mean percentage of low-salt diet is relatively small (1.7%), while percentage of low-sugar (regular) diet is around 2/3 (67.7%). The organic penetration rate in terms of spending is about 1% on average, which is consistent with the number documented in the literature. Households on average spend about 9.5% of their food budget on vegetables, 3.6% alcohol products, and less than 1% on tobacco products. The mean income for the pooled sample is $62,240, which is higher than the per-capita income in the United States between 2006 and 2008 ($53,881), as the households in the sample were from the metropolitan areas. The demographic profile of the sample is similar to the one from the census data. In the initial period of 26 weeks, the average weekly shopping frequency (number of shopping trips per week) is 1.66, equivalent to 7.2 shopping trips per month. On average, typical households spend around $20 in a single shopping trip, 1/3 of their products. The mean income for the pooled sample is $62,240, which is higher

than the per-capita income in the United States between 2006 and 2008 ($53,881), as the households in the sample were from the metropolitan areas. The demographic profile of the sample is similar to the one from the census data. In the initial period of 26 weeks, the average weekly shopping frequency (number of shopping trips per week) is 1.66, equivalent to 7.2 shopping trips per month. On average, typical households spend around $20 in a single shopping trip, 1/3 of their purchases are associated with discounts. Around 10% of households are value-oriented shoppers, who prefer low pricing and broad assortment and low service. The majority (90%) of households were shopping at traditional supermarkets, generally featuring promotional pricing, broad assortment in food categories and some service.

Table 4. Market Statistics, 23 U.S. Metropolitan Areas, 2006-2008

	Mean	Maximum	p75	Median	p25	Minimum
Population (1,000)	1,357.3	8,175.1	1,386.5	731.4	533.3	261.3
Market income ($10,000)	6.2255	7.8785	6.5411	6.1292	5.7205	5.0498
Market income inequality (Gini)	0.3371	0.3526	0.3417	0.3365	0.3307	0.3221
Market share (value-oriented)	16.99%	37.17%	22.84%	16.76%	10.82%	4.71%
Score (low salt-sugar-fat diet)	-0.0239	1.1019	0.1825	-0.0592	-0.2464	-0.8528
Organic penetration rate	1.43%	2.55%	1.72%	1.36%	1.05%	0.72%
Spending % on vegetables	9.47%	11.11%	10.34%	9.49%	8.79%	7.49%
Spending % on alcohol	3.65%	7.39%	5.41%	3.83%	1.67%	0.04%
Spending % on tobacco	0.81%	1.70%	1.07%	0.79%	0.46%	0.00%

Data source: Nielsen (2008) (author's calculations added).

Table 4 lists market-level statistics, including population, average income per capita, the degree of income inequality measured by Gini coefficient, the dietary styles (low salt-sugar-fat diet, organic penetration rate, spending percentage on vegetables, alcohol, and tobacco use), and market share of value-oriented retail format. There shows a vast range of differences among the metro markets. For instance, the biggest market is with five times of population compared to the smallest. Organic penetration rate is ranging from 0.72% to 2.55%. The market shares of value-oriented retail stores

also present a huge variation among markets. The value-oriented stores are known to concentrate in the southern and less-populated regions in the United States.

3.3 Results and Discussions

Table 5 contains the parameter estimates for household dietary choice from the two model specifications, the FE model and the market-factor model. The first equation (low salt-sugar-fat diet, labeled as "low ssf") is done by fixed effect or ordinary least square regression, while the rest four equations are estimated by Tobit with fixed effects or normal Tobit regression with left-censor at zero. Most estimates are statistically significant at 1% or 5%. The estimation results are discussed in the following four aspects: health awareness, ability to pay, price effect, shopping pattern, and market environment.

Health awareness is an important motive directing an individual's food purchase and dietary choice toward healthier lifestyles. Such awareness is highly linked to education and age groups. Generally speaking, more educated individuals are more aware of health information. It is indeed observed that college educated householders spend more in organic foods and vegetables, while consuming less tobacco products. Households with preschool children and elderly householders are more likely to adopt low salt-sugar-fat diet, and stay away from smoking, as suggested from the regression results. Households with school-age children, in general, are the families with middle-aged householders, who tend to less health aware compared to other age groups. Furthermore, school-age children are at their high-growing years, commonly eating fast foods or soft drinks. They tend to have high salt-sugar-fat diet and less vegetables. Alcohol and tobacco use are less likely to be observed in the households with children, preschool or school-age.

Individual's consumption choice is affected by ability to pay and prices. The higher-income has bigger purchasing power and is less sensitive to price changes. It explains why it is observed that households with higher income spend more on higher-

Table 5. Parameter estimates of household dietary choice

(1) The FE model

Variable	Low ssf	Organic	Vegetable	Alcohol	Tobacco
Household size	-0.0027	-0.0027**	-0.0014**	-0.0076**	0.0126**
Some college educated	-0.1448	0.0078**	0.0039**	0.0067**	-0.0316**
Married	-0.1004*	-0.0005	0.0098**	0.0036**	-0.0442**
Preschool children	0.2247**	0.0124**	-0.0014*	-0.0038*	-0.0345**
School-age children	-0.3162**	-0.0006	-0.0043**	-0.0046**	-0.0370**
Elderly	0.1558**	-0.0021**	0.0196**	0.0020	-0.0548**
Shopping frequency	0.0943**	0.0016**	-0.0067**	0.0141**	0.0376**
Average basket size	0.0023	0.0002**	-0.0004**	0.0010**	0.0026**
Discount use rate	-0.5802**	-0.0068**	0.0058**	0.0052**	-0.0640**
Value-oriented shopper	-0.3042**	-0.0079**	-0.0124**	-0.0321**	-0.0451**
Income	-0.0175**	0.0011**	0.0009**	0.0027**	-0.0067**
N	74,105	76,658	76,658	76,658	76,658
F statistics	78.83	86.01	8,898.03	254.69	173.27
(Prob > F)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)

(2) The market-factor model

Variable	Low ssf	Organic	Vegetable	Alcohol	Tobacco
Household size	0.0045	-0.0029**	-0.0016**	-0.0069**	0.0130**
Some college educated	-0.1612**	0.0083**	0.0025**	0.0166**	-0.0301**
Married	-0.1263**	-0.0003	0.0095**	0.0038**	-0.0444**
Preschool children	0.2353**	0.0123**	-0.0016*	-0.0034*	-0.0342**
School-age children	-0.3318**	-0.0005	-0.0039**	-0.0078**	-0.0383**
Elderly	0.1461**	-0.0021**	0.0189**	0.0071**	-0.0564**
Shopping frequency	0.1118**	0.0015**	-0.0065**	0.0132**	0.0379**
Average basket size	0.0034**	0.0002**	-0.0004**	0.0009**	0.0027**
Discount use rate	-0.6964**	-0.0063**	0.0050**	0.0030	-0.0665**
Value-oriented shopper	-0.308**	-0.0075**	-0.0127**	-0.0319**	-0.0453**
Income	-0.0175**	0.0011**	0.0009**	0.0026**	-0.0068**
Market average income	0.3118**	0.0032**	-0.0072**	0.0010	-0.0102**
Market income inequality	22.4088**	-0.0239	0.0960**	0.7405**	0.2217
Share(value-oriented)	-0.4591**	-0.0055*	0.0101**	0.0275**	0.1920**
Population (millions)	0.0946**	-0.0009**	0.0006**	-0.0032**	0.0013
Constant	-9.2311**	-0.0204*	0.1086**	-0.3162**	-0.2753**
N	74,105	76,658	76,658	76,658	76,658
F statistics	63.37	149.94	421.78	154.48	129.59
(Prob > F)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)

Data source: Nielsen (2008) (author's calculations added).

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

priced items, like organics, vegetables and alcohol, increases with household income, but purchase less tobacco products. The salt, sugar or fat nutrients generally do not yield price premiums over their alternatives. In fact, many items in this category, like low-fat milk, are priced even less than regular ones. This is likely the reason behind the observed result that the score of low-salt-sugar-fat diet decreases with household income. As to price sensitivity, those frequent discount users are more sensitive to price changes. Among the five dietary choices, prices of vegetables tend to fluctuate greatly due to changeable weather, storms, or natural disasters. It is consistent with the estimated result, in which the spending percentage on vegetables and discount use rate go in different directions.

The preference revealed by household's shopping pattern plays an influential role in dietary choice too. The frequent or big basket shoppers tend to purchase less vegetables but more organics, alcohol and tobacco items, compared with other shoppers. Interestingly, all five dietary styles, healthy or unhealthy, are negatively linked to value-oriented shopping loyalty. It appears that households pay more attention to the conventional food items, like meat, grain, dairy, than those specialty items considered in the estimation. After all, any of organics, vegetables, alcohol, or tobacco products consists less than 10% of total food spending on average as shown in the analysis data (see Table 3).

It is also examined whether market characteristics or locational attributes are influential to household dietary choice. Instead of market-specific fixed effects, four market-level indicators: market average income, market income inequality, market share of value-oriented, and population, are used to explain the variations between markets in the second model specification. It is suggested by the estimation results that adoption of low salt-sugar-fat diet or alcohol consumption is higher in the areas with high mean-income and high inequality, examples like New York, Chicago and Los Angeles. Consumers in the areas with low mean-income but high inequality (e.g. Tampa and some metro cities in South Atlantic division) tend to spend higher percentage of food expense on vegetables. In the areas where value-oriented retailers

have larger market shares, households tend to spend more percentage on vegetables, alcohol and tobacco, but less in organics, low salt-sugar-fat diet, compared to other areas in the country. Markets with more population tend to higher average adoption of low salt-sugar-fat diet and vegetables, but less in premium items (organics and alcohol).

Finally, the results of the regression estimation show significant predictors in almost all variables, but a surprising low R-squared value (unreported) in most equations of regression. Low R-squared values suggest that the model does not yield precise predictions on explaining the variations in dependent variables. This spells the complexity of the research question: people are fairly unpredictable, and dietary choice is not an exception. However, even when R-squared is low, low p-values for control variables still indicate a real relationship between the significant predictors and the response variable. Such associations are critical and what can be applied to gain understanding of complex behavioral decision making, like dietary choice in this study.

IV. Conclusion

This study uses a unique data set of the U.S. to examine the determinants of individual dietary choices and the linkages between the inequalities in diet, food consumption and income. Several key findings emerge from the analyses of 23 U.S. metro markets with over 70,000 of households in a three-year span from 2006 to 2008. First, it is shown that the degree of dispersion measured by Gini coefficient is generally greater in some health-related dietary pattern for organics, alcohol or tobacco use than in food consumption or in income. Provided its consequence to individual and public health, this result implies that the dietary inequality issue deserve further attention and discussion. Second, the dietary patterns vary greatly among markets and they have strong connections with income. Not only household income is influential to an individual's choice, but also the market-level income factors, especially income distribution of the market where households reside, are highly associated with

individual dietary decisions.

The study further examines the rationales behind household dietary choices through a comprehensive set of explanatory variables. The regression results provide supportive evidence that health awareness, ability to pay, price, shopping pattern and market environment are key to affect household dietary choice. With measures from household demographic characteristics, shopping behaviors, income, and market factors, this paper demonstrates how to map the observations from these variables shown in the food purchase data to surface out the underlying preference and motivation that drive household dietary choices. This study shows that aggregate level indicators, like average income, the degree of income inequality, market share information, can help us understand individual behavioral decision making in a meaningful way.

The analyses and discussions laid out in this study shall provide a great reference for literature and policy making in relevant to inequality, food consumption, health, and agriculture. Inequalities and dietary choice have important welfare meanings to all societies, not just the United States but also Taiwan or other countries. The study uses a large-scale and comprehensive transactional data to demonstrate how to decompose household food purchase records for further understanding of the association between dietary choice and income inequality. At the time of data collection, 2006-2008, there was no widespread use of smartphones and advanced scanning and tracking technologies. The household food purchases were collected through barcode scanners by participants at home. It is costly and may be inefficient, despite of its ingenious design. With current technologies, like A.I. (artificial intelligence) and IoT (internet of things), it is expected to seeing and having access to a data set recording comprehensive household food purchases, like the one used in this study, available in Taiwan or other countries in near future. With the approach and insights learned from the case of the U.S., policy makers can utilize the information of income distribution and its potential associations with individual health-related dietary choices to design relevant public health policy or nutritional subsidy program.

Endnotes

1. The maximum likelihood estimator under the setup of the Tobit model merges to a least-square estimator if the censored (zero) observations on y are close to none. This fits the vegetable equation in this study. Indeed, the estimation results of the vegetable equation from the Tobit regression are similar to the ones from the least square regression.
2. The households with less than 26 shopping records in the initialization period or with less than 52 records in the estimation period were excluded from the sample.
3. There were only 74,105 households included in the regression for low salt-sugar-fat diet equation due to missing values for some of households.

References

Allison, R. A., & Foster, J. E. (2004). Measuring health inequality using qualitative data. *Journal of Health Economics, 23*(3), 505-524.

Attanasio, O. P., Battistin, E., & Leicester, A. (2006). *From micro to macro, from poor to rich: Consumption and income in the UK and the US*. Working Paper, National Poverty Center, University College London. Retrieved from http://www.homepages.ucl.ac.uk/~uctpjrt/Files/Attanasio-Battistin-Leicester.pdf

Attanasio, O., Hurst, E., & Pistaferri, L. (2012). *The evolution of income, consumption, and leisure inequality in the US, 1980-2010*. (National Bureau of Economic Research, Working Paper No. w17982). Retrieved from https://www.nber.org/chapters/c12675.pdf

Baumgärtner, S., Drupp, M. A., Meya, J., Munz, M., & Quaas, M. F. (2012). *Income distribution and willingness to pay for ecosystem services*. Paper presented at the 19th Annual Conference of the European Association of Environmental and Resoiurces Economists. Retrieved from http://bioecon-network.org/pages/13th_2011/Baumgaertner.pdf

Bell, D. R., Ho, T., & Tang, C. S. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research, 35*(3), 352-369.

Berry, S. T., & Haile, P. A. (2009). *Nonparametric identification of multinomial choice demand models with heterogeneous consumers*. (National Bureau of Economic Research, Working Paper No. w15276). Retrieved from https://www.nber.org/papers/w15276.pdf

Blundell, R., & Preston, I. (1998). Consumption inequality and income uncertainty. *Quarterly Journal of Economics, 113*(2), 603-640.

Blundell, R., Pistaferri, L., & Preston, I. (2008). Consumption inequality and partial insurance. *The American Economic Review, 98*(5), 1887-1921.

Boumtje, P. I., Huang, C. L., Lee, J. Y., & Lin, B. H. (2005). Dietary habits, demographics, and the development of overweight and obesity among children in the United States. *Food Policy, 30*(2), 115-128.

Bureau of Labor Statistics (2016). *The Consumer Expenditure Survey*, the Bureau of Labor Statistics, U.S. Department of Labor. Retrieved from https://www.bls.gov/cex/home.htm.

Chang, H. H., & Nayga, R. M. (2009). Television viewing, fast-food consumption, and children's obesity. *Contemporary Economic Policy, 27*(3), 293-307.

Dougkas, A., Reynolds, C. K., Givens, I. D., Elwood, P. C., & Minihane, A. M. (2011). Associations between dairy consumption and body weight: A review of the evidence and underlying mechanisms. *Nutrition Research Reviews, 24*(1), 72-95.

Dubois, P., Griffith, R., & Nevo, A. (2014). Do prices and attributes explain international differences in food purchases? *The American Economic Review, 104*(3), 832-867.

Erreygers, G., & Van Ourti, T. (2011). Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. *Journal of Health Economics, 30*, 685-694.

Fisher, J. D., Johnson, D. S., & Smeeding, T. M. (2013). Measuring the trends in inequality of individuals and families: Income and consumption. *The American Economic Review, 103*(3), 184-188.

Furst, T., Connors, M., Bisogni, C. A., Sobal, J., & Falk, L. W. (1996). Food choice: A conceptual model of the process. *Appetite, 26*(3), 247-266.

Golan, E., Kuchler, F., Mitchell, L., Greene, C., & Jessup, A. (2001). Economics of food labeling. *Journal of Consumer Policy, 24*(2), 117-184.

Hassett, K. A., & Mathur, A. (2012). A new measure of consumption inequality. *AEI Economic Studies, 2*, 1-22.

Hsieh, M.-F. (2012). Estimating price effect on consumer shopping across quality-differentiated store formats. *Taiwanese Agricultural Economic Review, 18*(1), 1-30.

Hsieh, M.-F., Mitchell, P. D., & Stiegert, K. W. (2009). Potato demand in an increasingly organic marketplace. *Agribusiness: An International Journal, 25*(3), 369-394.

Hsieh, M.-F., & Stiegert, K. W. (2012). Store format choice in organic food consumption. *American Journal of Agricultural Economics, 94*(2), 307-313.

Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science, 21*(3), 229-250.

Kreuter, M. W., & Wray, R. J. (2003). Tailored and targeted health communication: Strategies for enhancing information relevance. *American Journal of Health Behavior, 27*(1), S227-S232.

Krueger, D., & Perri, F. (2006). Does income inequality lead to consumption inequality? Evidence and theory. *The Review of Economic Studies, 73*(1), 163-193.

Lin, B. H., & Guthrie, J. F. (2012). *Nutritional quality of food prepared at home and away from home*, 1977-2008. (Economic Information Bulletin No. 105, Economic Research Service, United States Department of Agriculture). Retrieved from https://ageconsearch.umn.edu/record/142361/

Lin, B. H., Huang, C. L., & French, S. A. (2004). Factors associated with women's and children's body mass indices by income status. *International Journal of Obesity, 28*(4), 536-542.

Madden, D. (2010). Ordinal and cardinal measures of health inequality: An empirical comparison. *Health Economics, 19*(2), 243-250.

Meyer, B. D., & Sullivan, J. X. (2009). *Five decades of consumption and income poverty*. (National Bureau of Economic Research, Working Paper 14827). Retrieved from https://www.nber.org/papers/w14827.pdf

Meyer, B. D., & Sullivan, J. X. (2010). *Consumption and income inequality in the U.S.: 1972-2005*. (Working Paper, University of Notre Dame). Retrieved from https://economics.stanford.edu/sites/g/files/sbiybj9386/f/inequality7.4a_with_tables.pdf

Naga, R. H. A., & Yalcin, T. (2008). Inequality measurement for ordered response health data. *Journal of Health Economics, 27*(6), 1614-1625.

Naga, R. H. A., & Yalcin, T. (2010). *Median independent inequality orderings*. Discussion Paper No. 2010-118, Scottish Institute for Research in Economics (SIRE). Retrieved from http://sticerd.lse.ac.uk/dps/darp/darp103.pdf

Nielsen, Inc. (2008). *Nielsen Homescan Data, ERS-USDA version, 2006-2008*. Economic Research Service, U.S. Department of Agriculture. Retrieved from http://www.nielsen.com/clients/index.html.

Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (1990). Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries. *The American Economic Review, 80*(5), 1139-1156.

Reutterer, T., & Teller, C. (2009). Store format choice and shopping trip types. *International Journal of Retail and Distribution Management, 37*(8), 695-710.

Saarela, A. M. (2014). Change of behaviour when selecting food products in a supermarket environment after reminding consumers about weight management. *Public Health Nutrition, 17*(5), 1147-1155.

Song, I., & Chintagunta, P. K. (2007). A discrete-continuous model for multicategory purchase behavior of households. *Journal of Marketing Research, 44*(4), 595-612.

Strauss, J., & Thomas, D. (1998). Health, nutrition, and economic development. *Journal of Economic Literature, 36*(2), 766-817.

Tang, C. S., Bell, D. R., & Ho, T.-H. (2001). Store choice and shopping behavior: How price format works. *California Management Review, 43*(2), 56-74.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26*, 24-36.

Van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: Validation of a new approach to measurement. *Journal of Health Economics, 22*(1), 61-87.

Van Strien, T., Herman, C. P., & Verheijden, M. W. (2009). Eating style, overeating, and overweight in a representative Dutch sample. Does external eating play a role? *Appetite, 52*(2), 380-387.

Ver Ploeg, M. L., & Ralston, K. L. (2008). *Food stamps and obesity: What do we know?* (Economic Information Bulletin No. 34, Economic Research Service, United States Department of Agriculture). Retrieved from https://www.ers.usda.gov/webdocs/publications/44221/12196_eib34_1_.pdf

Verbeke, W. (2008). Impact of communication on consumers' food choices. *Proceedings of the Nutrition Society, 67*(3), 281-288.

Winters, P., Essam, T., Zezza, A., Davis, B., & Carletto, C. (2010). Patterns of rural development: A cross country comparison using microeconomic data. *Journal of Agricultural Economics, 61*(3), 628-651.

Yen, S. T., Lin, B. H., & Davis, C. G. (2008). Consumer knowledge and meat consumption at home and away from home. *Food Policy, 33*(6), 631-639.

飲食選擇與所得不均：運用家庭食品購買資料之實證研究*

謝銘逢**

分配不均是資本主義的特徵及重要的社會問題。本文針對所得不均與家戶的飲食選擇之關聯進行研究，深入比較美國主要的都會城市的飲食模式差異、以及與食品消費和所得分配之間的關聯性。本實證研究使用一個涵蓋數個市場的美國家戶食品購買資料，來檢驗社經和人口因素、購物習性及所得變數對健康相關的飲食選擇，如低鹽、低糖、低脂飲食、有機食品、蔬菜、酒及香煙草等之影響。本研究之迴歸結果顯示，對健康的意識度、支付能力、價格、購物習性和市場環境等因素，是影響個人飲食選擇的關鍵。不僅個人所得對其飲食選擇有顯著影響，總體層面的所得分配，包含每人平均所得和所得不均度，也有關鍵解釋能力。這些結果幫助我們進一步了解複雜的飲食選擇決策，及其與所得分配的關聯。

所得不均和飲食選擇在各個社會都有重要的福利意義，不僅是本研究所分析的美國，對於台灣或其他國家均具參考價值。本研究利用巨量家戶實際購買交易資料，提供了一個分析家戶食品購買資料的方法，以從中了解飲食選擇與所得不均的關聯。這些結論可以提供公共政策的決策參考，利用所得分配的資訊及其與健康相關的飲食選擇之關聯，用以設計相關的公共衛生政策或營養補貼計畫。

關鍵詞：飲食選擇、所得分配、不均度、食品消費、家戶購物決策
***JEL Codes**: I14*

---

* 本研究承科技部專題研究計畫經費補助（103-2410-H-002-026/106-2410-H-029-010），謹致謝忱。本文承蒙主編及兩位匿名審查人所提供的寶貴評論及建議，特此致謝。

** 東海大學經濟學系助理教授。

投稿日期：2020 年 3 月 30 日；第一次修改日期：2020 年 5 月 1 日；第二次修改日期：2020 年 5 月 7 日；接受日期：2020 年 5 月 21 日。