Shop Around and You Pay More*

Ji Yan[a,c], Kun Tian[b,c,d], Huw Dixon[b,d], Saeed Heravi[b,d], Peter Morgan[b,d]

[a] *Durham University. UK*
[b] *Cardiff University, UK*
[c] *Xiangtan University, China*
[d] *Centre for Price and Inflation at Cardiff, UK*

Abstract

We examine the relationship between the prices paid by households and their shopping patterns measured in terms of shopping frequency and the range of stores visited. We use the TNS data which allows us to control for household heterogeneity. The main contribution of the paper is that we find proper instruments to correct for endogeneity of shopping patterns. And we find that there exists a robust positive relationship between the price paid and the number of store visited. We argue that visiting larger number of stores makes households less likely to save using store-loyalty discounts.

Key Words

Shopping frequency, price, store loyalty, demographics

* This paper is previously circulated in title of "Life-cycle prices and shopping frequency". We are grateful for their useful comments and suggestions to Erwan Gautier and participants in Academy of Marketing 2013, Royal Economic Society 2014 annual conference. The authors thank TNS UK Ltd. for providing the panel and survey data used in this research. The use of TNS UK Ltd data in this work does not imply the endorsement of TNS UK Ltd. in relation to the interpretation or analysis of the data. All errors and omissions remain the responsibility of the authors.

1. Introduction

The relationship between shopping frequency and the price households pay for items has been of interest both to economists (Aguiar and Hurst, 2005; 2007; Kaplan and Menzio, 2014) and marketing scholars and practitioners (Ainslie and Rossi, 1998; Kim and Park, 1997; Ma et al., 2011). However, the relationship is not straightforward: the decisions of how often to shop, the range of stores to visit and the products to purchase are all linked to underlying variables reflecting the preferences and constraints faced by households. In this paper we adopt an empirical methodology that enables us to allow for the endogeneity of consumers' shopping pattern and control for heterogeneity in respect of consumer's characteristics so we can discover the underlying relationships. We carry out our study using the Taylor Nelson Sofres (TNS) home scan dataset, which contains over 30 million transaction records including detailed information about expenditure, quantity, brand, size, store, etc. There are more than 11,000 households from 10 regions across the UK over the period 2002-2005.

We start by constructing household-specific price indices. For the purpose of comparison, we follow the approach adopted by Aguiar and Hurst (2007) to construct a household-specific relative price index, which is defined as a ratio of the actual expenditure of household to its hypothetical expenditure at average prices. We also construct a price index based on a household utility function. This utility-based price index can account for household's preference heterogeneity as well as allowing for households to respond optimally to price variations. We can then compare the utility based price index based on the actual expenditure of each household relative to the hypothetical at average prices. The AH relative price index is basically a weighted arithmetic mean, whilst the utility based index is a geometric mean.

We first investigate the dispersion of both price indices across different age groups. Our first finding is that people in their later fifties and above pay higher prices than their younger counterparts, with or without controls for shopping needs and household characteristics. We then investigate how shopping frequency varies across different age groups. We find that older shoppers make more shopping trips per month than younger shoppers. Thus older people shop more frequently but end up paying more for their groceries. Our finding is in sharp contrast to those in Aguiar and Hurst (2007) [2], who find that in the US older people pay less by shopping more frequently. To understand better the relationship between shopping frequency and price paid, we decompose the shopping frequency into two

[2] Kaplan and Menzio (2014) get the same results as Aguiar and Hurst (2007) by using the same data source but with larger number of observations and longer sampling period.

components: i) number of store visited, and ii) number of trips to each store. We find that households across different age groups (and demographics) have different shopping patterns. For example, older households visit more stores and make more trips to each store than their younger counterparts.

The most important contribution of this paper is that we test the causal relationship between shopping frequency and price. There is a superficial correlation reflected in the ordinary least squares (OLS) estimates which suggests a negative relation between shopping frequency and price (we find this in our UK data as do Aguiar and Hurst (2007) and Kaplan and Menzio (2014) for similar US data). To account for the endogeneity of shopping patterns, we carefully select a set of instruments and estimate the relationship between shopping patterns and price paid using instrumental variables (IV). We also estimate a simultaneous equations system by 3SLS method to allow for linkages between equations by common shocks (SURE). Both IV and 3SLS regression results suggest that the higher number of stores visited leads to a higher price paid. Doubling the number of stores visited will increase the AH relative price index by 5%. Moreover, there is a negative relationship between number of store visited and the savings in expenditure by discount usage – households who visit more stores benefit less from discounts for example. Thus, after allowing for endogeneity and consumer heterogeneity we find that shopping around for groceries means that you end up paying more for items on average.

The theoretical literature in economics and marketing have highlighted a range of possibilities about how these three variables (shopping frequency, number of stores and the price paid) are linked. Optimal search theory would indicate that a rational household facing search costs would visit as many stores as possible to obtain the best price for each item. Households with lower search costs would search more and thus obtain lower prices. Thus, since time is a key element in search costs, households with more plentiful time (the retired and unemployed for example) might be expected to visit more stores and obtain lower prices. Kaplan and Menzio (2014) find for US scanner data that there is a negative relationship between the number of stores visited and the average price paid. However, the behavioural dimension is also stressed by the marketing literature: a higher shopping frequency increases the likelihood of "spontaneous purchases" that account for up to half of a consumer's purchase and a large scale of sales for both retailers and manufacturers (Block and Morwitz, 1999; Ho et al., 1998). “Spontaneous purchases” is seen as impulse buying behaviour that results in unplanned purchases(Stilley et al., 2010a). Unplanned purchases allow consumers to try new products and alternative brands to those regularly purchased, and result in increasing consumers' total budget and more time spent in decision-making (Bell et al., 2011; Inman et al., 2009; Stilley et al., 2010b). More importantly, recent marketing evidence has found that unplanned purchases are unlikely to be triggered by promotions (Hui et al., 2013). Hui et al. (2013) use portable vision tracking system on shoppers to observe their field of vision in store and thus record their product consideration and purchase conversion. Their findings indicate that although promotions are effective in catching a shopper's eyes, they have little effect on increasing the likelihood of putting the promotional products into the shopper's basket. Hence, unplanned purchases, which account on average for more than half of a consumer's purchase, are not generally purchased under promotion. Moreover, most unplanned purchase are done by consumers who have less time pressure and/or tend to visit more numbers of stores (Bucklin and Lattin, 1991; Hui et al., 2013; Park et al., 1989). Hence, consumers who spend more time in shopping and visiting more stores are more likely to conduct unplanned purchases. Although unplanned purchase behaviour is not the focus of this study, it reveals the possibility that behavioural factors may explain why more frequent shopping does not result in lower prices paid. Our empirical results would seem to indicate that the behavioural considerations probably outweigh the rational economic factors. This is not a small effect: we found that consumers' store variety seeking behaviour enlarges the price differences across household types by up to 30%.

The paper also shows how shopping patterns in terms of shopping frequency and numbers of stores visited vary across demographics and explores who pays higher prices. These patterns are important for retailer to infer consumers' demographic characteristics from shopping frequency and target particular group of consumers who are willing to pay higher price. We find store variety seeking levels that differ across different household type based on UK home scan datasets, which contrasts with US datasets (Aguiar and Hurst, 2007) where no such store variety seeking behaviour was found.

The rest of the paper is organized as following. Section 2 describes the dataset and builds household-specific price index. In Section 3, we investigate the price index dispersion and the variation of shopping patterns across age groups. The relationship among price-paid, shopping patterns, and discounts usage is estimated by using instrumental variable and 3SLS respectively. In Section 4, we check the robustness of findings in Section 3 by running regressions across different household types. In Section 5, we conclude.

2. Data

2.1 TNS data

Our analysis uses Taylor Nelson Sofres (TNS) data[4] in Great Britain. The data set includes information about an entire range of Fast-Moving Consumer Goods (FMCG) for 156 weeks from Oct 2002 to Dec 2005. The household purchase dataset is designed to record all consumer goods purchased by the household at a variety of stores. The household's purchase details are recorded by panellists who use hand held barcode scanners, which is functionally based on the Universal Product Code (UPC) that records a wealth of product and brand sales information and information such as the category of goods, the brand name, size, flavour, etc. The purchasing data from each household are transferred directly to TNS's central computer. Non-bar-coded products purchases are recorded according to the shopping receipts mailed from households. Our dataset comprises of 21 categories products as main product group, 15 for food (beer, wine, spirits, fish, meat, vegetables, fruit, egg, dairy products, sugar, bread, butter, oil, soft drink and tea) and 6 for non-food (including pet food, cleaning material, medical, personal care products and small tools). Within 21 main groups of products, we have 189 product categories and 185,495 brands in total. Within our TNS data, the average weekly grocery expenditure is £37 per week. The average weekly expenditure in our dataset is similar to £44 per week from ONS's report in 2005, therefore, the dataset covers majority of daily life shopping needs on grocery and non-grocery expenditures.

We use the TNS dataset across all 10 regions in the UK. The households were regionally balanced in order to represent the household population equivalently. We have 10,000+ UK households and all household demographics information is collected from ONS's survey data and is updated continuously. Specifically, the biographical information contains: the main shopper's age, marital status, social class, gender, children number, home ownership, car ownership, the number of toilets in a house, the employment status and family size, etc. We categorize households according to the age of main shopper into 9 cohorts: 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, and 65-74. The households with age below 25 and above 75 are excluded from our study, and they only account for 6% of the transaction observations in the whole sample. We also distinguish the households by a mixed-indicator, which reflects the household's marriage, children and age characteristics. The indicators are: senior single, senior couple, single adult, couple no child, other no child, couple with child, other with child, lone parent. Descriptive statistics of household variables

[4] Griffith et al. (2009), Griffith and O'Connell (2009, 2010), Leicester and Oldfield (2009), Griffith and Nesheim (2013) also use the TNS Homescan data. TNS UK is a part of the Kantar, a large market research company. The TNS data are part of Kantar Worldpanel data set.

are provided in Table 1. The dataset enable us to examine the variation of the price paid across different demographic characteristics. However, because the data is cross-sectional in nature and we must be aware that some of our results may be confounded with cohort effects.

Table 1. Household Age and Types

Household Age Range	*Observations*	*Share*
25-29	598	5.28
30-34	1,403	12.39
35-39	1,575	13.91
40-44	1,300	11.48
45-49	1,206	10.65
50-54	1,215	10.73
55-59	1,255	11.08
60-64	1,026	9.06
65-74	1,746	15.42

Household Type	*Observations*	*Share*
Lone parent	287	2.53
Other with children	456	4.03
Couple with children	3,281	28.97
Other, no children	3,399	30.02
Couple, no children	364	3.21
Single adult	765	6.76
Senior couple	2,570	22.7
Senior single	202	1.78

Household Marital status	*Observations*	*Share*
Married	7,492	66.16
Single/ Widowed/ Divorced/ Separated	3,832	33.84

Number of toilets in house	*Observations*	*Share*
0	9	0.08
1	7,001	61.82
2	3,520	31.08
3	712	6.29
4	67	0.59
5	11	0.1
6	3	0.03
10	1	0.01

Household Region	*Observations*	*Share*
Anglia	784	6.92
Lancashire	1,403	12.39
London	2,192	19.36
Midlands	1,880	16.6
North East	572	5.05
Scotland	968	8.55
South	1,055	9.32
South West	355	3.13
Wales and West	991	8.75
Yorkshire	1,124	9.93
Total	11,324	100

Household Size	*Observations*	*Share*
1	2,013	17.78
2	3,821	33.74
3	1,908	16.85
4	2,385	21.06
5	857	7.57
6	254	2.24
7	62	0.55
8	19	0.17
9	2	0.02
10	2	0.02
11	1	0.01

Moreover, purchase records are validated appropriately to maintain reliability and consistency of the data. For each purchase, the data provides information about a precise account of: panel ID number (household indicator), the product category, brand specification, UPC code, the shopping trip by date, the store visited, grand weight of the item purchased, package size, price per pack, the number of packs bought, the amount of money spent on each purchase (measured by British Sterling), the amount of money saved by discounts. Therefore, we can work out the price paid by each household each year and also the money saved by using discounts per household per year.

Within the TNS dataset, we have 12,477 households and 39,883,691 observations in total. We study households in which the average age of "main shopper" is no less than 25 and no more than 75. It leaves us 11,324 households and 37,334,775 observations. The dataset also contains transaction information from 222 stores, including grocery stores, convenient stores, specialty stores and price-cutting stores.We provide an overview of four categories of stores and four big chain stores in UK. Among the big four chain stores, the percentage of chain expenditures across all households at Chain-1 is 32.77, Chain-2 is 29.65, Chain-3 is 26.65, and Chain-4 is 10.93. Expenditure shares are computed for each household-year pair and averaged over four big grocery store chains. As a result, the expenditure shares for the big four chain stores sum to one.

The advantage of our data is that it includes biographical information about the consumers making the purchases, and it records the transactions across a variety of retail stores. Since the TNS dataset has the information of each shopping trips and store visited, we can work out the shopping frequency and store variety[5] directly.We define the monthly shopping frequency as number of shopping trips by household in a given month. We denote the number of shopping trips conducted by household $h \in H$ during month $t \in T$ by $freq_{h,t}$. The total shopping trips can be decomposed into two parts: number of store visited and number of trips per store. Hence, we have the identity that $freq_{h,t} = ns_{h,t} \times tps_{h,t}$. Here $ns_{h,t}$ and $tps_{h,t}$ represent number of store visited by household h during month t, and the number of trips for each identical store respectively.

2.2 Household-specific price index

[5]Store variety is defined as the number of stores visited during a given period. Consequently, we can show that the higher the store variety is, the lower the store loyalty is.

We first build a household-specific price index following the same approach as in Aguiar and Hurst (2007)[7], we build a price index that is used to compare the actual cost of a household's shopping basket to the cost of the identical shopping basket at average prices. Specifically, we denote the expenditure of good $k \in K$ shopped by household $h \in H$ on shopping trip $l \in t$ t by E_{khl}, and the relevant quantity bought by q_{khl}. And monthly total quantity of good k bought by household h is $q_{kht} = \sum_{l \in t} q_{khl}$. Accordingly, we can calculate the actual expenditure of household h on good k during month t as $E_{k,h,t} = \sum_{l \in t} E_{k,h.,l}$. Similarly, household h's actual total expenditure in month t is $E_{h,t} = \sum_{k \in K} E_{k,h,t}$ We can define the average price of good k as a ratio between the total expenditure across different households and the total quantity bought by all households, which can be shown as:

$$\bar{p}_{k,t} = \frac{\sum_{h \in H} E_{k,h,t}}{\sum_{h \in H} \sum_{l \in t} q_{k,h,l}} \qquad (1)$$

Then, we can calculate the cost of a basket of goods under average prices as:

$$\bar{E}_{h,t} = \sum_{k \in K} \bar{p}_{k,t} q_{k,h,t} \qquad (2)$$

Therefore, the relative price index for the household h during month t can be calculated as the ratio of expenditures at actual prices divided by the cost of the basket at the average price:

$$\tilde{P}_{h,t} = \frac{E_{h,t}}{\bar{E}_{h,t}}$$

We normalise the relative price index for household h by dividing through the average relative price index across households:

$$P_{h,t} = \frac{\tilde{P}_{ht}}{1/N_t \sum_{h \in H} \tilde{P}_{ht}} \qquad (3)$$

where N_t is the number of households in month t. The price index constructed above is similar to Laspeyres index that the basket of goods is the same though the prices between numerator and denominator can vary. We need to point out that differences in our price index just reflect price differentials for the identical goods, but do not reflect differences in the quality of goods purchased.

[7] Kaplan and Menzio (2014) adopt the same approach as in Aguiar and Hurst (2007).

The average relative price index used by Aguiar and Hurst (2007) is based on a simple expenditure weighted arithmetic average of the prices paid. However, if we are interested in household utility, then consumer theory indicates that we should focus on the cost of utility using a price index. A price index can reflect the fact that consumers can substitute away from more expensive items and towards cheaper items. In order to do this, we assume that households have a simple Cobb-Douglas utility function[8] which has the useful property that the expenditure shares we observe in the data can be directly linked to the underlying preference parameters (the exponents of the direct utility function). The household's cost of living (price index) is then defined as the geometric average of the prices paid with weights given by the expenditure shares (multiplied by a term which also depends on the same expenditure shares). Specifically, we build a Cobb-Douglas utility based price index as following: At time t household h consumes a subset of goods $k(h,t) \subseteq K$. Total expenditure for household h at time t is $E_{h,t}$. Expenditure for good k by household h at time t is $E_{k,h,t}$. Hence the share of expenditure on good k by household h is $e_{k,h,t} = E_{k,h,t} / E_{h,t}$. The direct utility conditional on the set of goods is:

$$U_{h,t} = U_{h,t}(\mathbf{q}) = \prod_{k \in K(h,t)} \left(q_{k,h,t}\right)^{e_{k,h,t}}$$

The corresponding price index is the cost of one unit of utility:

$$P_{h,t} = \left[\prod_{k \in K(h,t)} \left(e_{k,h,t}\right)^{e_{k,h,t}} \right] \prod_{k \in K(h,t)} \left(p_{k,h,t}\right)^{e_{k.h.t}}$$

The term in square brackets is often ignored, because it is a “constant”. However, here it will vary. Taking logs gives us:

$$\log P_{h,t} = \sum_{k \in K(h,t)} e_{k,h,t} \log p_{k,h,t} + \left[\sum_{k \in K(h,t)} e_{k,h,t} \log e_{k,h,t} \right]$$

in terms of observables $E_{k,h,t}$ and $q_{k,h,t}$ these are:

$$P_{h,t} = \left[\prod_{k \in K(h,t)} \left(e_{k,h,t}\right)^{e_{k,h,t}} \right] \prod_{k \in K(h,t)} \left(\frac{E_{k,h,t}}{q_{k,h,t}} \right)^{e_{k,h,t}}$$

$$\log P_{h,t} = \sum_{k \in K(h,t)} e_{k,h,t} \log E_{k,h,t} - \sum_{k \in K(h,t)} e_{k,h,t} \log q_{k,h,t} + \left[\sum_{k \in K(h,t)} e_{k,h,t} \log e_{k,h,t} \right]$$

[8] We could of course use more general forms of utility function, such as CES preferences. However, this would require the specification of the elasticity of substitution for each household which would be hard to do without being arbitrary.

We also normalise the household utility based price index by dividing through the average utility base price index across households.

3. Average Price Paid and Shopping Patterns across Age Groups

In this section, we document the price index dispersion across age groups. We then describe how shopping patterns vary across age groups. In addition to comparing our results to those in Aguiar and Hurst (2007)from the US homescan data, we estimate a simultaneous system of equations that can account for endogeneity and correlated shocks . For the purpose of concreteness, we first describe the price index dispersion for a specific product, with a specific brand, in a specific size.

3.1 An example: Heinz Baked Bean 4x415 ml

We take Heinz Baked Bean 4x415 ml as an illustrative example. For this specific product (Baked Bean) with specific brand (Heinz) and specific size (4x415 ml), we plot the average relative price index across 9 age-range groups in Figure 1. We normalise the mean value of logarithm of relative price index for age 25-29 to zero. Similarly, we document household's shopping patterns for Heinz Baked Bean 4x415 ml. Accordingly, we plot the mean shopping frequency and mean number of store visited.

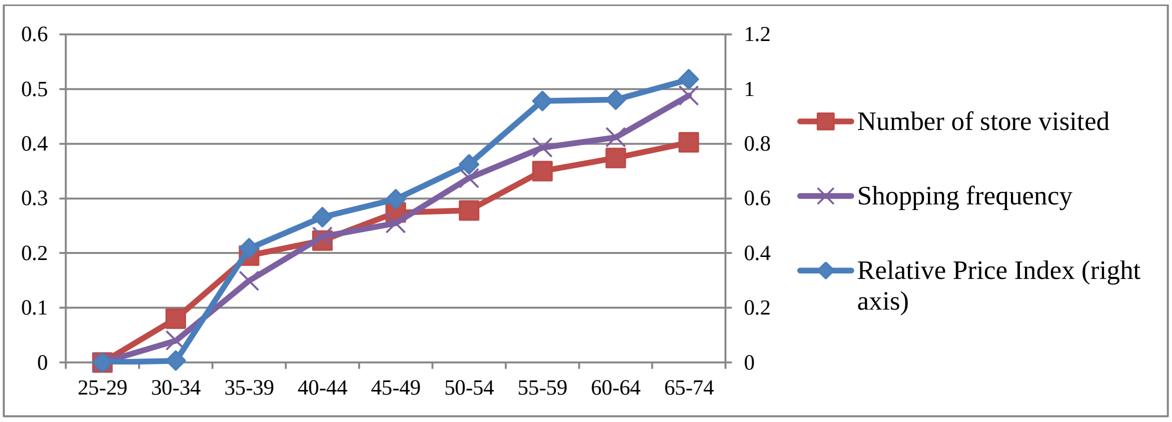

Figure 1. Normalised mean value of relative price index, number of store visited and shopping frequency across age ranges for all transactions relating to Heinz Baked Bean 4x415ml.

Figure 1 clearly depicts that average relative price index for Heinz Baked Bean 4x415 ml increases with age. Meanwhile, the shopping frequency and number of store visited also increases with age. To see whether this is an exceptional or representative case, we next investigate price-paid and shopping frequency across age groups for all the goods in our dataset.

3.2 Life Cycle Price

In this section, we compare the price paid among households in different age groups. We estimate the following equation by OLS:

$$\log P_{h,t} = \beta' A_{h,t} + \gamma' SN_{h,t} + \eta' HC_{h,t} + \lambda_{r(h)} + \kappa_t + \varepsilon_{h,t} \qquad (4)$$

$\log P_{h,t}$ is the logarithm of the price index for household h in month t. $A_{h,t}$ represents age-range dummy variables, dividing the whole sample into 9 cohorts; $SN_{h,t}$ represents household's shopping needs measured by a set of variables , including quantity of goods purchased, number of brands purchased, and the number of product categories purchased; $HC_{h,t}$ is a set of household-specific characteristic, which include employment status dummy, household size; $\lambda_{r(h)}$ is the dummy for region where household lives, and κ_t is the year dummy. $\varepsilon_{h,t}$ and $\nu_{h,t}$ represent residual in each regression equation respectively.

Table 2. Price Paid and Shopping Frequency across Age Groups
Log Deviation from Age 25-29

	Dependent Variable			
	ln(Relative Price Index)		ln(Utility Price Index)	
Regressors	**I**	**II**	**III**	**IV**
Age 30-34	-0.001	-0.000	0.050**	0.005
	(0.001)	(0.001)	(0.019)	(0.011)
Age 35-39	0.002	0.004*	0.142***	0.002
	(0.001)	(0.001)	(0.019)	(0.011)
Age 40-44	-0.000	0.002	0.177***	0.023
	(0.001)	(0.001)	(0.021)	(0.012)
Age 45-49	0.001	0.002	0.188***	0.007
	(0.002)	(0.001)	(0.021)	(0.013)
Age 50-54	0.004*	0.003*	0.208***	0.052***
	(0.001)	(0.001)	(0.022)	(0.014)
Age 55-59	0.010***	0.008***	0.275***	0.097***
	(0.001)	(0.001)	(0.023)	(0.015)
Age 60-64	0.012***	0.010***	0.224***	0.079***
	(0.001)	(0.001)	(0.023)	(0.016)
Age 65-74	0.015***	0.011***	0.209***	0.082***
	(0.001)	(0.001)	(0.020)	(0.014)
Include Controls for Shopping Needs and Household Characteristics	No	Yes	No	Yes
N	38,782	38,782	38,782	38,782

* p<0.05, ** p<0.01, *** p<0.001

Notes: Columns I - IV reports coefficients on age range dummies when dependent variable is the average log price index, without and with controls for shopping needs and household characteristics (including household size, employment status), respectively. Regions dummies and year dummies are also controlled but not reported. Results in Columns I and II are based on relative price index generated by Aguiar and Hurst (2007). Results in Columns III and IV are based on price index generated by Cobb-Douglas Utility function. Column V and VI report results of regressing average log shopping frequency on age range dummies, without and with controls for shopping needs and household characteristics, respectively. Numbers in parentheses indicate robust standard errors which are clustered at the household level. N represents number of household-year pair.

The regression results of Equation (4) are presented in Table 2 and Figure 2. Column I of Table 2 reports the results of a regression of logarithm of relative price index on age-range dummies for 38,782 household-year observations. Column I of Table 2 is the regression counterpart of diamond-line in Figure 2. The results indicate that households in their young and middle age stage pay roughly the same price for the constant basket of goods. However, households in their late fifties pay rather higher price. For example comparing to households in their late forties, households in their late fifties pay prices that are 0.8 per cent higher while households in their early sixties and late sixties/early seventies pay prices that are 1 per cent and 1.2 per cent higher respectively. Moreover, these differences are all statistically significant. Column II of Table 2 is the regression counterpart of square-line in Figure 2, reporting the results of a regression of logarithm of relative price index on age-range dummies given the household's shopping needs, characteristics, as well as region and year dummies been controlled. The coefficients on the age dummies indicate that the inclusion of these controls provides similar results

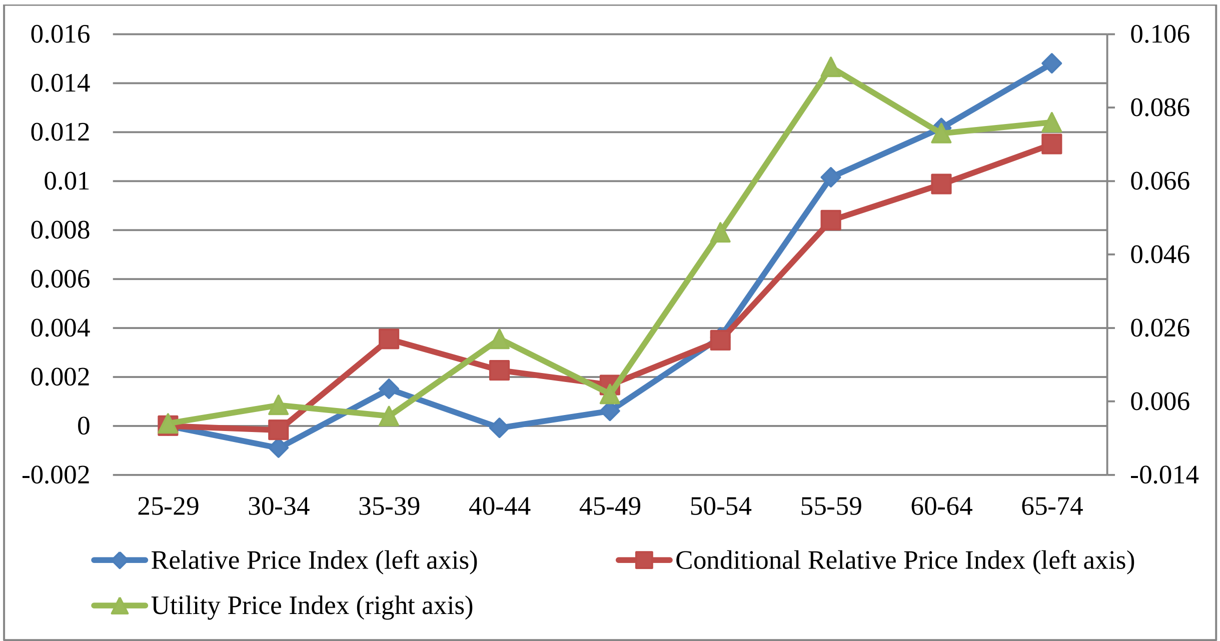

Figure 2. Relative Price Index and Utility Price Index across Age Ranges

Using utility based price index, we get similar estimated results with average price index increasing with age . Column III and column IV of Table 2 are regression of logarithm of utility based price index on age range dummies without or with other control variables respectively. Column IV of Table 2 is the regression counterpart of pyramid-line in Figure 2. The green pyramid-line shows that households in the age range of 60-64 pay about 8 per cent higher prices than those younger than 49, given that shopping needs and household characteristics are controlled. We expect that households substitute away from expensive goods to cheaper goods in the utility based price index. This means that the benefits of lower prices for specific goods are greater than in the relative price index. Conversely, the cost of paying higher prices is greater. Given the price index dispersion they face, the 60-64s would like to find some lower prices. The arithmetic average assumes that they would continue to purchase exactly the same basket at the new lower prices. The utility based index allows them to reallocate their expenditure to take advantage of the cheaper goods. The utility based index shows that the gains and losses are much larger than implied by the relative price index.
As is clear from Figure 1, the price index increases with age. It is in sharp contrast to the finding from US home scan data in Aguiar and Hurst (2007), and Kaplan and Menzio (2014). However, it is similar to the finding of Abe and Shiotani (2014) by investigating Japanese home scan data.

3.3 Life Cycle Shopping Pattern

Since we find that older households pay higher price than their younger counterpart, we would like to see how shopping patterns vary across age groups. We estimate OLS regression

$$\log freq_{h,t} = \theta' A_{h,t} + \rho' SN_{h,t} + \varphi' HC_{h,t} + \lambda_{r(h)} + \kappa_t + \nu_{h,t} \qquad (5)$$

$$\log ns_{h,t} = \tau' A_{h,t} + \iota' SN_{h,t} + \pi' HC_{h,t} + \lambda_{r(h)} + \kappa_t + \nu_{h,t} \qquad (6)$$

$$\log tps_{h,t} = \zeta' A_{h,t} + \xi' SN_{h,t} + \chi' HC_{h,t} + \lambda_{r(h)} + \kappa_t + \nu_{h,t} \qquad (7)$$

$\log freq_{h,t}$ is the logarithm of shopping frequency for household h in month t. $\log ns_{h,t}$ is the logarithm of number of store visited for household h in month t. $\log tps_{h,t}$ is the logarithm of number of trips per store for household h in month t. The remaining variables are defined the same as previous subsection.

Table 3. Number of Store Visited and Trips per Store Visited across Age Groups
Log Deviation from Age 25-29

	Dependent Variable					
	ln(Shopping Frequency)		ln(Num.of Store Visited)		ln(Trips per Store visited)	
Regressors	**I**	**II**	**III**	**IV**	**V**	**VI**
Age 30-34	0.041	0.010	0.020	0.006	0.022	0.004
	(0.032)	(0.028)	(0.026)	(0.026)	(0.015)	(0.012)
Age 35-39	0.146***	0.037	0.078**	0.028	0.068***	0.010
	(0.031)	(0.027)	(0.026)	(0.025)	(0.015)	(0.012)
Age 40-44	0.249***	0.126***	0.146***	0.091***	0.103***	0.034**
	(0.033)	(0.028)	(0.027)	(0.027)	(0.016)	(0.012)
Age 45-49	0.340***	0.166***	0.175***	0.102***	0.165***	0.064***
	(0.033)	(0.029)	(0.028)	(0.027)	(0.016)	(0.012)
Age 50-54	0.474***	0.275***	0.261***	0.177***	0.213***	0.098***
	(0.033)	(0.029)	(0.028)	(0.027)	(0.016)	(0.013)
Age 55-59	0.601***	0.305***	0.375***	0.237***	0.226***	0.068***
	(0.033)	(0.028)	(0.028)	(0.027)	(0.016)	(0.012)
Age 60-64	0.631***	0.319***	0.379***	0.233***	0.253***	0.086***
	(0.034)	(0.030)	(0.029)	(0.029)	(0.016)	(0.014)
Age 65-74	0.812***	0.429***	0.512***	0.320***	0.300***	0.109***
	(0.031)	(0.029)	(0.026)	(0.027)	(0.015)	(0.013)
Include Controls for Shopping Needs and Household Characteristics	No	Yes	No	Yes	No	Yes
N	38,782	38,782	38,782	38,782	38,782	38,782

* p<0.05, ** p<0.01, *** p<0.001

Notes: Column I and II presents the results of regressing log number of store visited per month on age dummies without and with controls for shopping needs and household characteristics (including household size, employment status), respectively. Regions dummies and year dummies are also controlled but not reported. Column III and IV presents the results of regressing log number of trips per store per month on age dummies without and with controls for shopping needs and household characteristics, respectively. Our data sample includes households between the ages of 25 and 74, inclusive. The bench mark age group in each regression is 25-29. For households aged 25-29, the mean number of stores visited is 2.7, and the mean number of trips per store is 1.9. Numbers in parentheses indicate robust standard errors which are clustered at the household level.

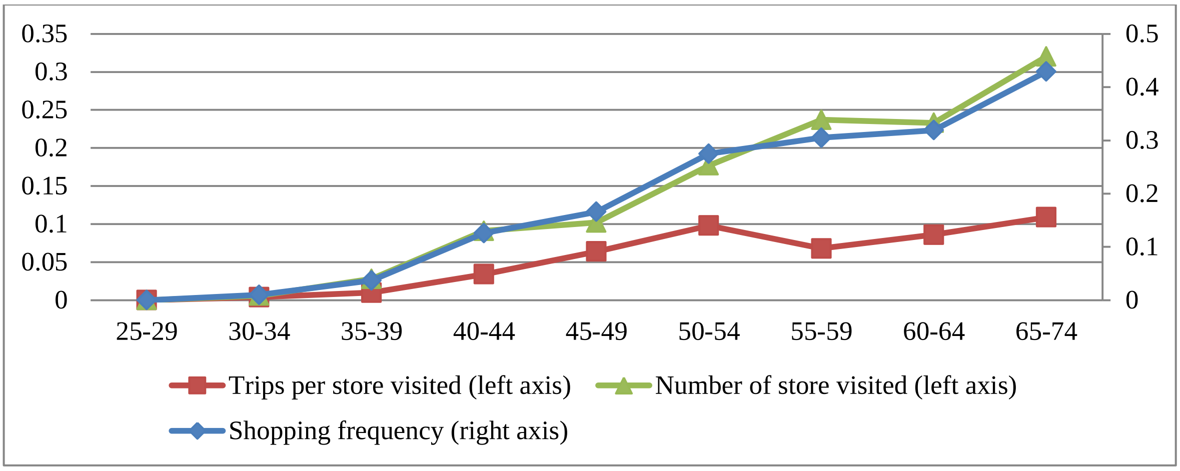

Figure 3. Shopping Frequency, Number of Store Visited and Trips per Store Visited across Age Ranges

Table 3 and Figure 3 report the regression results of shopping patterns (shopping frequency, number of store, and trips per store) on age-range dummies without/with controlling for shopping needs and household characteristics. Again, controls for shopping needs and household characteristics do not significantly change the life-cycle shopping patterns. Column II of Table 3 is the regression counterpart of diamond-line in Figure 3. For reference, shoppers aged 25-29 make 5.1 shopping trips in each month on average. The figure indicates that shoppers in their early part of life cycle (younger than 39) remain roughly the same shopping frequency. However, older shoppers make more shopping trips per month than younger shoppers. Specifically, comparing to shoppers in their late forties, shoppers in early fifties are about 10 per cent higher in shopping frequency. Relative to shopper's aged 45-49, shoppers aged 55-59, 60-64, and 65-74 are 14 per cent, 16 per cent, and 26 per cent higher in shopping frequency respectively. The regression results suggest that the old people tend to do shopping more frequently, even when we control the shopping needs and household characteristics.

As discussed in previous section, the shopping frequency can be decomposed into two components: number of stores visited per month, and number of trips made per store each month. We measure shopper's store variety by counting the average number of stores visited in a given month. Besides, store intensity is measured by calculating the average number of trips made by shoppers for each store in a given month.

In Figure 3, we plot the two component measures across different age range. We construct two series in the same way as for overall shopping frequency. Specifically, the store variety (number of store visited) and the store intensity (number of trips per store) are presented as log deviations from the benchmark group, the household with age range 25-29. Furthermore, the shopping needs and household characteristics across different age group are controlled for the purpose of meaningful comparison. On average, the benchmark group “25-29” age shoppers visit 2.7 different stores per month and make 1.9 trips for each store per month. Figure 3 indicates that older shoppers visit more different stores than the younger counterparts. Specifically, the households aged 65-74 visit about 20 per cent more different stores than the households aged 45-49.

Moreover, Figure 3 also shows that the older shoppers shop more frequently for each store than their younger counterpart. For example, the households aged 65-74 visit each store about 5 per cent more frequently than the households aged 45-49. Accordingly, the increase in overall shopping frequency by older shoppers reflects more shops visited and more frequent visits to each store.

Our findings are different to those in Aguiar and Hurst (2007), which claims that “older shoppers visit the same number of stores as their younger counterparts. The increase in total shopping trips by older households reflects more frequent visits to each store”. More number of stores visited reflects lower “store loyalty” and higher “store variety”. Hence, the low “store loyalty” shoppers may less likely to take advantage of in-store promotion attached with store loyalty cards. Furthermore, shoppers visiting more stores are more likely to spread expenditures across stores and with higher likelihood in purchasing in higher priced retailer. Above all, the difference in store variety accounts for the major part of difference in overall shopping frequency. We expect that this key feature of shopping pattern have a dominant effect on price paid during life cycle.

3.4 Elasticity of Price with respect to Shopping Patterns

In this subsection, we estimate the elasticity of price to the variation in shopping frequency, number of stores visited per month, and trips per store visited. We regress households-specific price indices (relative price index and utility based price index) on aspects of households’ shopping behaviour. The results are reported in Table 4 for relative price index and Table 5 for utility based price respectively. For comparison reason, we focus on Table 4. Columns I-IV reports OLS regression results, showing that relative price indices are lower both for households who shop more frequently and for households who visit more stores. Our OLS regression results are in line with those findings in Aguiar and Hurst (2007) and Kaplan and Menzio (2014). However, a natural question is whether the OLS results are reliable. To answer this question, we have to examine the basic assumption for OLS that the independent variables are exogenous.

It is natural that the price change will influence people's shopping patterns including the shopping frequency and store variety seeking. In other words, there is potential "reverse" causation in our study, and OLS suffers from "endogeneity problem". In this situation, the OLS regression generally produces biased and inconsistent estimates. We conduct Durbin-Wu-Hausman tests to see whether shopping frequency/number of store visited is exogenous in our study. The test results reject the null hypothesis that shopping frequency/number of store visited is exogenous. Therefore, we consider using some instrumental variables to correct for endogeneity. A valid instrument is a variable that is uncorrelated with the statistical residual term and is correlated with the endogenous explanatory variables, conditional on the other covariates. Specifically, we use a variable known as an instrument, which affects the dependent variable (price paid) only through the endogenous regressor (ShoppingFrequency/Number of store visited). If the assumption held, instrumental variable allows us to use variation in shopping frequency/store variety to identify its effect on price, while ignoring variation in price correlated with omitted factors.

One potential instrument variable for shopping frequency, following Aguiar and Hurst (2007), is the age range of main shoppers. This supports our assumption that the value of shopping time varies across life cycle. The F-statistics at first stage regression show that age dummies are correlated with endogenous regressors. However, the over identification test (e.g. Sargan test) suggests that the age range dummies fail to satisfy exclusion restriction and are not valid IVs[9].

One possible source of endogeneity is omitted variable due to unobserved preference in shopping. People may different in their attitude to shopping. For example, someone may

[9] We replicate Aguiar and Hurst (2007) results by using their data, and find that age dummies as instruments cannot pass over identification test.

Table 4. Effect of shopping patterns on household relative price index

	I	II	III	IV	V	VI	VII	VIII
Shopping frequency	-0.004***				0.015			
	(0.000)				(0.009)			
Number of store visited		-0.006***		-0.006***		0.026*		0.062***
		(0.000)		(0.000)		(0.011)		(0.011)
Trips per store visited			0.002*	0.001			-0.015	-0.082***
			(0.001)	(0.001)			(0.009)	(0.015)
Regression type	OLS	OLS	OLS	OLS	IV	IV	IV	IV
Instrument set					Number of toilets dummies	Number of toilets dummies	Number of toilets dummies	Marital and Number of toilets dummies
Minimum eig.val.					89.79	36.41	77.86	4.20
N	38782	38782	38782	38782	38782	38782	38782	38782

* p<0.05, ** p<0.01, *** p<0.001

Notes: Column I-XI presents the results of 2SLS estimation with controls for employment status, log number of household size, household shopping needs (household average log number of brands purchased per month, the average log number of product categories, and log number of quantity purchased), home ownership, social class, regions dummies and year dummies. Numbers in parentheses indicate standard errors.

Table 5. Effect of shopping patterns on household utility price index

	I	II	III	IV	V	VI	VII	VIII
Shopping frequency	0.071***				0.125**			
	(0.005)				(0.044)			
Number of store visited		0.049***		0.059***		0.339***		0.195*
		(0.005)		(0.005)		(0.083)		(0.089)
Trips per store visited			0.107***	0.125***			0.034	-0.192
			(0.012)	(0.012)			(0.098)	(0.138)
Regression type	OLS	OLS	OLS	OLS	IV	IV	IV	IV
Instrument set					Number of toilets dummies	Number of toilets dummies	Number of toilets dummies	Marital and Number of Toilets dummies
Minimum eig.val.					92.24	35.56	74.74	3.79
N	38782	38782	38782	38782	38782		38782	

* p<0.05, ** p<0.01, *** p<0.001

Notes: Column I-XI presents the results of 2SLS estimation with controls for shoppers' age, squared value of age, employment status, number of times shopping by car, coupon usage, log number of household size, squared log number of household size, household shopping needs (household average log number of brands purchased per month, the average log number of product categories, and log number of quantity purchased), home ownership, social class, regions dummies and year dummies. Numbers in parentheses indicate standard errors.

enjoy the shopping process itself rather than searching for lower price. Given the nature of our dataset, we use toilet number in a house as a proxy for this difference in preference.

Another possible source of endogeneity is to consider if shopping patterns affect price paid via opportunity cost of time. We choose the marital status of main shoppers as an indicator of the opportunity cost of time. The premise here is that a married shopper has more housework, spend less time in shopping and would like one-stop shopping pattern.

We reports regression results with respect to relative price index using instrumental variables in Column V-VIII in Table 4. In Column V-VII, we use number of toilets as instrument. In Column VIII, we use both marital status and number of toilets as instruments. The Sagan tests suggest that the number of toilet in a house and marital status variable satisfy exclusion restriction and both are valid instruments. As shown in Column VIII, the estimation of elasticity of price to number of stores is 0.043, which suggests that doubling the number of stores visited will raise the price paid by 4.3 per cent.

We also report effect of shopping patterns on household utility based price index in Table 5. We find that OLS and 2SLS estimations show consistent results that higher shopping frequency and more stores visited lead to higher price paid.

Since we observe a positive correlation between household price index and number of stores visited in single equation 2SLS estimation, we would like to test this relationship within a simultaneous system of equations. For the identification issue, we only include two equations in the system (the price equation, and the number of store equation). Table 6 reports 3SLS system equation estimation results. Specifically, System 1 and System 2 only differs in price index specification. System equation results are quite robust under different price index setting: Increasing in number of store visited will increase the price paid and older households visit more stores. We need to point out that the coefficient on "Age" in price equations (Column I or Column III of Table 6) has different interpretation to that on age-range dummies in Table 2. In Column I or Column III of Table 6, the negative coefficient on Age means that given shopping patterns and other factors the same, older households tend to pay less. However, in Table 2, the coefficient on age-range dummies reflect price paid variation due to the potential change in shopping patterns across age-ranges. What interests us is how variation of shopping patterns across age groups affects the price paid by each age group, rather than unexplained age effect on price paid.

Table 6. Hypotheses testing results with an endogenous system of equations based on Model

	I System 1 Ln (Relative Price Index)	II System 1 Ln(Number of Store Visited)	III System 2 Ln (Utility Price Index)	IV System 2 Ln(Number of Store Visited)
Ln(Number of Store	0.049***		0.458***	
Visited)	(0.003)		(0.030)	
	-0.000	0.007***	-0.001***	0.007***
Age	(0.000)	(0.000)	(0.000)	(0.000)
	0.004***	-0.067***	0.016**	-0.065***
Employment	(0.001)	(0.007)	(0.005)	(0.007)
	0.014***	-0.189***	0.036***	-0.189***
Shop by car	(0.001)	(0.004)	(0.007)	(0.004)
Coupon usage	-0.068***	-0.527***	-0.476***	-0.528***
propensity	(0.003)	(0.034)	(0.030)	(0.034)
		-0.073***		-0.197***
Ln(household size)		(0.013)		(0.014)
Chi-2	2723.97	13818.73	36543.10	13811.13
p-value	0.000	0.000	0.000	0.000
Number of				
Observation	38776	38776	38776	38776

* p<0.05, ** p<0.01, *** p<0.001
Notes: This table reports the results from a system of equations and controls for shoppers' age, employment status, number of times shopping by car, coupon usage, log number of household size, and squared log number of household size. Other control variables such as household shopping needs (household average log number of brands purchased per month, the average log number of product categories, and log number of quantity purchased), home ownership, social class, regions dummies and year dummies are also controlled but not reported. Number of toilets is used as instrument variable. Numbers in parentheses indicate standard errors.

3.5 Life-Cycle Discount Usage

We find in previous subsection that older households pay higher price while visit more stores than their younger counterpart. This raises the additional question of how a shopper pay higher prices by visiting higher number of stores. One possible explanation is that visiting more stores makes the shopper less likely to use store-loyalty discount. Store loyalty schemes range from giving coupons to allowing significant cash discounts for staying with the same store for the weekly shop. According to a recent report in the Sunday Times, "Younger people are more likely to use loyalty cards and 'buy into' retailers' promotional strategies, and so are more likely to get a better deal." Hence, we investigate how the share of expenditure saved by discounts varies over life cycle.

Figure 4 plots the share of expenditures saved by discounts over life cycle. As before, this result is conditional on household's characteristics and shopping needs. Again, we let 25-29 be as benchmark group. The value for the other age-range groups reflects logarithm deviations from households age 25-29. It can be seen from the Figure 4 that the share of expenditures saved by discounts is roughly constant through the age of 49. Households age 65-74, however, the share of expenditures saved by discounts is lower than over 1 per cent (p-value<0.01), all else equal.

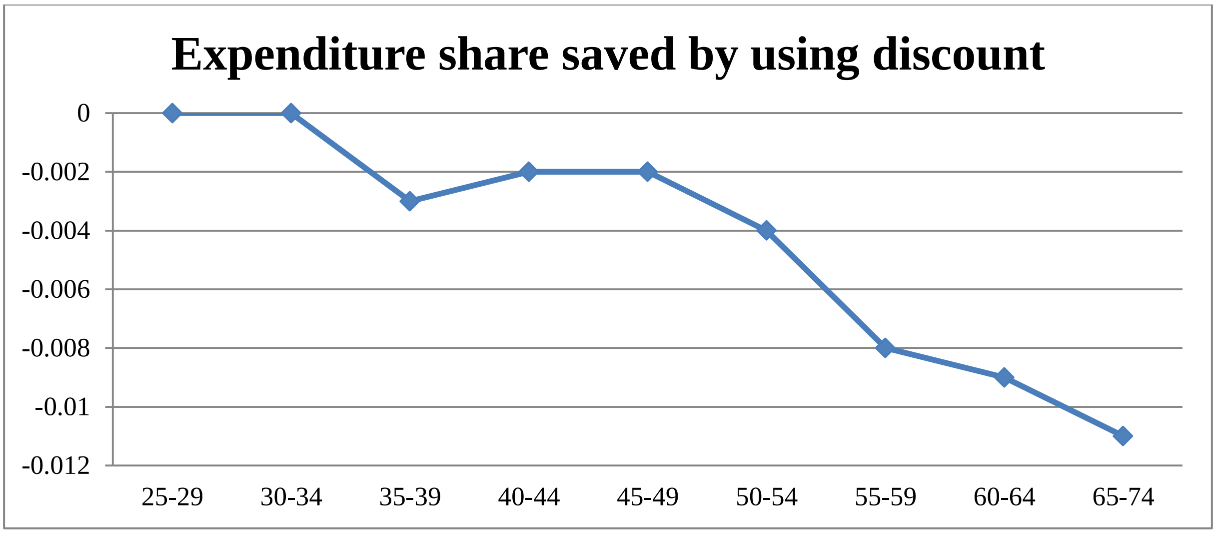

Figure 4. Expenditure share saved by using discount across age ranges

We regress the logarithm of share of expenditure saved using discounts on the logarithm of number of stores visited, controlling for shopping needs and household characteristics. We use the number of toilets as instrument variable to correct endogenous bias. The main result is that doubling the number of store visited is associated with a 3.6 percentage point decrease in the share of expenditure saved by discounts.

4. Robustness Check: Price Paid and Shopping Patterns across Household Types

In this section, we look at the average price paid and the shopping patterns for each household type. Our purpose is to check whether the positive relationship between price paid and the number of store visited still hold across different household type.

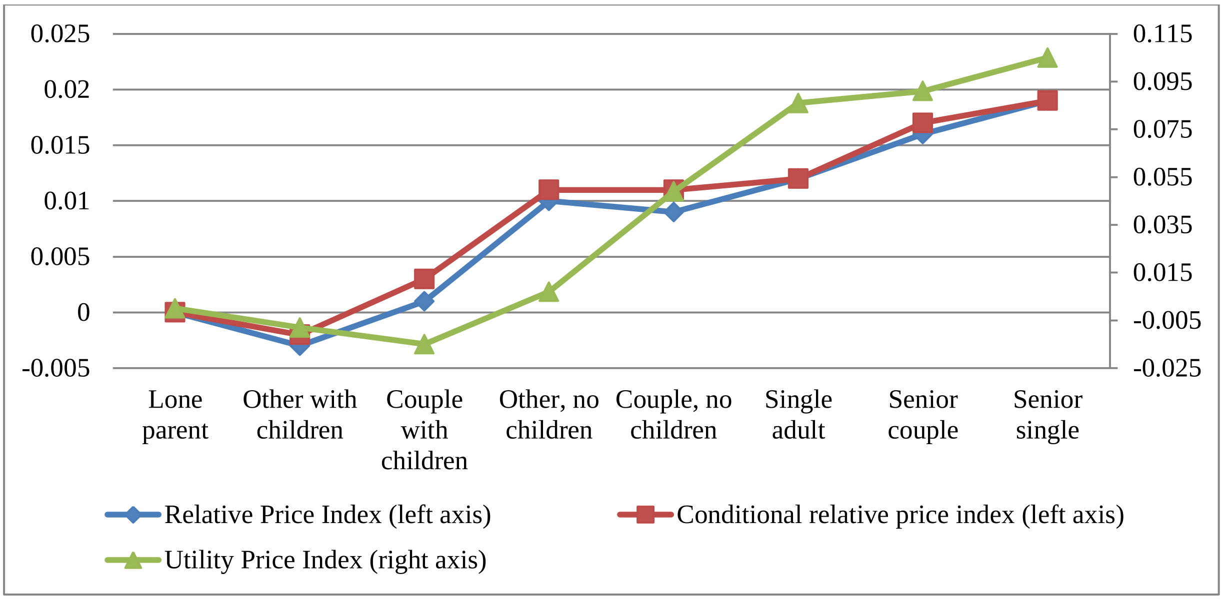

Figure 5. Relative Price Index and Utility Price Index across Household Types

Similar to previous section, we plot price paid and shopping patterns across household respectively. In Figure 5, the square-line shows the relative price index varies across household types after controlling for shopping needs and household characteristics. Specifically, we find that relative to those “lone parent” households (the baseline group), the “single adult” households pay 1.2 per cent higher price. The “couple no child” households pay about 1.1 per cent higher price than the “couple with child” households under the relative price index. Utility price index shows similar pattern across household types.

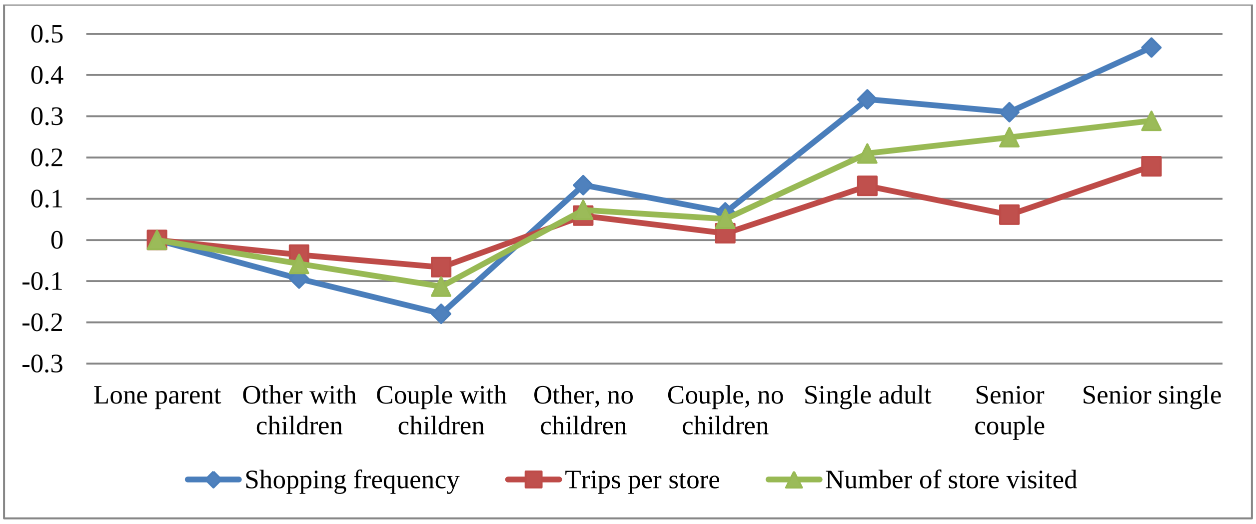

Figure 6. Shopping Frequency, Number of Store Visited and Trips per Store Visited across Household Types

In Figure 6, we see that the "couple no child" households shop 24 per cent more frequently than the "couple with child". The "single adult" households shop 31 per cent more frequently than the "lone parent". The households with children shop less often than those without children. In Figure 6, we also plot number of store visited and trips per store in log deviation from benchmark group across different household types, with shopping needs being controlled. The benchmark group is "lone parent", which on average visits 3.1 stores each month and shops 2.0 times for each store. Figure 6 indicates that households without child shop in more stores than households with child. Specifically, the "single adult" households visit 18.4 per cent more shops than the "lone parent" households. Moreover, "couple without child" shop in 12.1 per cent more stores than the counterpart "couple with child". However, the store intensity (number of trips per store) differences between comparable groups are quite small. For example, the "couple with child" households only shop 6 per cent less frequently per store than the counterpart "couple without child". Hence, the difference in overall shopping frequency across household types is largely drawn from the difference in store variety rather than store intensity across household types.

We re-estimate 3SLS system equations, replacing "Age" variable by household type dummies. The main estimation results show that the higher the number of store visited, the higher the price paid. It suggests that the positive relationship between store variety and price paid is quite robust.

We finally plot the expenditure share saved using discounts across household types. As shown in Figure 7, it is clear that childless households save less by using discount than their counterpart.

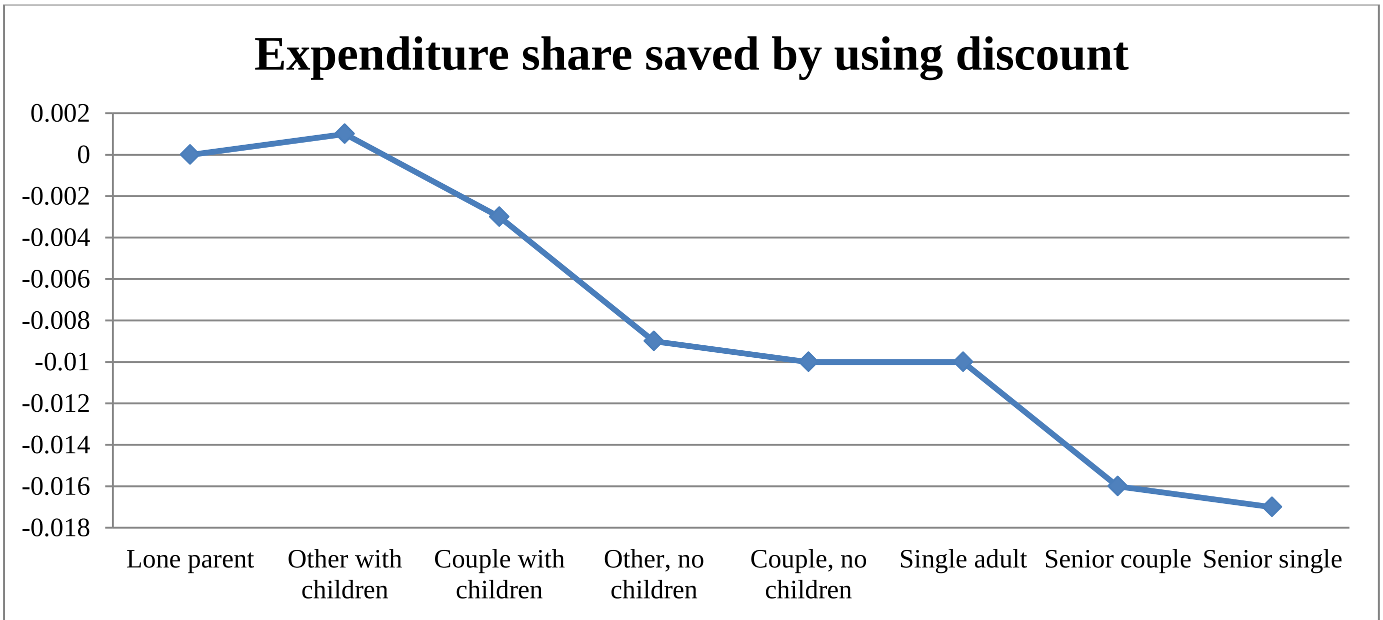

Figure 7. Expenditure share saved by using discount

5. Conclusion

This study provide a comprehensive analysis of the impact of the shopping patterns on the prices paid across households; it also generalizes the findings of how shopping frequency and store variety seeking behaviour vary across an entire set of FMCGs. Most consumer studies focused on grocery purchases only, non-food purchases receive intermittent attention of researchers mainly due to unavailability of the data set. Our analysis covers not only food but also non-food datasets helps us to document consumers' shopping patterns more comprehensively. Recent study based on food and non-food purchases documents the potential and actual saving that consumers realize from different shopping choices(Griffith et al., 2009), and our study extends it in terms of analysing how willing to pay more, instead of how much, they pay for the same basket of goods.

Based on 30+ millions UK scanner data points, we find that relations between shopping patterns and price paid across demographics are completely different from what was found in the US. In our study, the older and consumers without children visit more stores and pay more. In the US , the older consumers who pay lower prices shop more often, visit the same store more often but not visit more numbers of stores(Aguiar & Hurst, 2007). This is important information for UK retailers because it is critical to realize how big differences exist across countries, and it is helpful for them to revaluate the importance of attracting existing customers to visit more frequently.

Scanner data on household consumption are rich in information such as UPC (barcode) for specific products which differs in terms of manufactures and reputations, the sizes and chains about stores, and biographical information of households. Further research may take consumers' inventory and transportation costs into consideration and develop a comprehensive demand analysis in order to discover how and to what extend households are able adopt their shopping behaviour in terms of selecting what to buy, when to buy, how much to buy and where to buy.

References

Abe, N. and K. Shiotani (2005). Who faces higher prices? An empirical analysis based on Japanese Homescan data', Asian Economic Policy Review, vol. 9(1), pp. 94-115.

Aguiar, M. and E. Hurst (2005). 'Consumption versus expenditure', Journal of Political Economy, vol. 113(5), pp. 919-948.

Aguiar, M. and E. Hurst (2007). 'Life-cycle prices and production', American Economic Review, vol. 97(5), pp. 1533-1559.

Ainslie, A. and P. E. Rossi (1998). 'Similarities in choice behavior across product categories', Marketing Science, vol. 17(2), pp. 91-106.

Bell, D. R., D. Corsten and G. Knox (2011). 'From Point of Purchase to Path to Purchase: How Preshopping Factors Drive Unplanned Buying', Journal of Marketing, vol. 75(1), pp. 31-45.

Bell, D. R. and J. M. Lattin (1998). 'Shopping behavior and consumer preference for store price format: Why "large basket" shoppers prefer EDLP', Marketing Science, vol. 17(1), pp. 66-88.

Block, L. G. and V. G. Morwitz (1999). 'Shopping Lists as an External Memory Aid for Grocery Shopping: Influences on List Writing and List Fulfillment', Journal of Consumer Psychology, vol. 8(4), pp. 343-375.

Bucklin, R. E. and J. M. Lattin (1991). 'A 2-State Model of Purchase Incidence and Brand Choice', Marketing Science, vol. 10(1), pp. 24-39.

Dubois, P., R. C. Griffin and A. Nevo (2014). 'Do Prices and Attributes Explain International Differences in Food Purchases?', American Economic Review, vol. 104(3), pp. 832-867.

Griffith, R., E. Leibtag, A. Leicester and A. Nevo (2009). 'Consumer Shopping Behavior: How Much Do Consumers Save?', Journal of Economic Perspectives, vol. 23(2), pp. 99-120.

Griffith, R. and L. Nesheim (2013). 'Hedonic methods for baskets of goods', Economics Letters, vol. 120(2), pp. 284-287.

Griffith, R. and M. O'Connell (2009). 'The Use of Scanner Data for Research into Nutrition', Fiscal Studies, vol. 30(3-4), pp. 339-365.

Griffith, R. and M. O'Connell (2010). 'Public Policy towards Food Consumption', Fiscal Studies, vol. 31(4), pp. 481-507.

Ho, T. H., C. S. Tang and D. R. Bell (1998). 'Rational shopping behavior and the option value of variable pricing', Management Science, vol. 44(12), pp. S145-S160.

Hui, S. K., Y. L. Huang, J. Suher and J. J. Inman (2013). 'Deconstructing the "First Moment of Truth": Understanding Unplanned Consideration and Purchase Conversion Using In-Store Video Tracking', Journal of Marketing Research, vol. 50(4), pp. 445-462.

Inman, J. J., R. S. Winer and R. Ferraro (2009). 'The Interplay Among Category Characteristics, Customer Characteristics, and Customer Activities on In-Store Decision Making', Journal of Marketing, vol. 73(5), pp. 19-29.

Kaplan, G. and G. Menzio (2014). 'The Morphology and Price Dispersion', working paper, The National Bureau of Economic Research.

Kim, B. D. and K. Park (1997). 'Studying patterns of consumer's grocery shopping trip', Journal of Retailing, vol. 73(4), pp. 501-517.

Leicester, A. and Z. Oldfield (2009). 'Using Scanner Technology to Collect Expenditure Data', Fiscal Studies, vol. 30(3-4), pp. 309-337.

Ma, Y., K. L. Ailawadi, D. K. Gauri and D. Grewal (2011). 'An Empirical Investigation of the Impact of Gasoline Prices on Grocery Shopping Behavior', Journal of Marketing, vol. 75(2), pp. 18-35.

Park, C. W., E. S. Iyer and D. C. Smith (1989). 'The Effects of Situational Factors on in-Store Grocery Shopping Behavior - the Role of Store Environment and Time Available for Shopping', Journal of Consumer Research, vol. 15(4), pp. 422-433.

Stilley, K. M., J. J. Inman and K. L. Wakefield (2010). 'Planning to Make Unplanned Purchases? The Role of In-Store Slack in Budget Deviation', Journal of Consumer Research, vol. 37(2), pp. 264-278.

Stilley, K. M., J. J. Inman and K. L. Wakefield (2010). 'Spending on the Fly: Mental Budgets, Promotions, and Spending Behavior', Journal of Marketing, vol. 74(3), pp. 34-47.