

# Selective Bargain Hunting. A Concise Test of Rational Consumer Search\*

Chloe Lee

Wen Long

María José Luengo-Prado

UC Berkeley

Barclays Capital

Federal Reserve Bank of Boston

Bent E. Sørensen

University of Houston and CEPR

December 13, 2019

## Abstract

A model of time-allocation between work, leisure, and price-search for different goods predicts that rational consumers should spend relatively more effort searching for better prices of preferred goods of which they consume relatively more. Using scanner data, we confirm empirically the implication that consumers find better prices for their preferred goods.

**JEL classification:** E21, D12

**Keywords:** Consumption; Consumer Search; Time Use

---

\*Lee: UC Berkeley, School of Information, 102 South Hall # 4600, Berkeley, CA 94720-4600, email: chloe.hy.lee@berkeley.edu; Long: Barclays Capital, 745 7th Ave, New York, NY, 10019, email: longwen1983@gmail.com; Luengo-Prado: Federal Reserve Bank of Boston, 600 Atlantic Ave., Boston, MA, 02210, email: maria.luengo-prado@bos.frb.org. Sørensen: Department of Economics, University of Houston, 204 McElhinney Hall, Houston, TX 77204, email: besorensen@uh.edu. The content is solely the responsibility of the authors and does not represent the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

## 1 Introduction

A rational consumer trades off the value of time and effort spent on finding lower prices with the value of time and effort spent on other pursuits. We conduct a simple novel test of rational bargain hunting/price-search: A consumer who frequently drinks soda but seldom drinks beer should rationally spend more effort finding low prices for Coke than finding low prices for Budweiser. Using detailed shopping information from the IRI Academic Dataset, we confirm this prediction. Our results are relevant for understanding consumer rationality: The finding that consumers search more for lower prices of some goods than others provides direct evidence that consumers face constraints on time and/or attention, and that they rationally allocate these scarce resources to optimize welfare. Our findings are also relevant for optimal store pricing and for understanding how consumers can efficiently self-insure by strategically finding lower prices following employment or retirement shocks.

The IRI dataset records the purchases made by a panel of households over an 11-year period at a selection of stores in Eau Claire, Wisconsin, and Pittsfield, Massachusetts. The prices are recorded for items at the UPC (scanner-code) level, which constitutes the finest possible definition of a good. For each transaction, both the price of the item and the quantity purchased are reported, and we relate the price a consumer pays for each item in a given month to the average price paid by all consumers during the same month for the same item. We show that the price paid on average correlates negatively with the amount purchased during a given month.

Stigler (1961), in a pioneering paper, suggested that information is scarce and consumers invest time in finding favorable prices—an activity that he labeled “search.” As summarized in Kaplan and Menzio (2013), many recent papers examine price-search using scanner data under the heading of “bargain hunting.” Aguiar et al. (2013) use the American Time Use Survey to show that households in states with higher unemployment spend relatively more time on home production and shopping, and Coibion et al. (2015) use scanner data from IRI to show that consumers obtain better prices during recessions by switching to cheaper stores.<sup>1</sup> Nevo and Wong (2019), using Nielsen Homescan data, show that during the Great Recession consumers obtained lower prices by, among other practices, using more coupons, purchasing more items on sale, and shopping more often at “big box” stores.<sup>2</sup> Nevo and Wong (2019) also find that the return to shopping declined during the Great Recession, so the increased amount of search is consistent with a lower shadow value of time.

The literature has found intuitively reasonable differences in shopping behavior across individuals. Aguiar and Hurst (2007) show that retirees spend relatively more time shopping, and Stroebel and Vavra (2014), using changes in house prices to isolate exogenous changes in wealth, find that wealthier households spend relatively less time shopping. Chevalier and Kashyap (2019) examine purchases using the IRI data and posit a model with two types of consumers: (1) “shoppers,” who pay the best price possible because they chase discounts, substitute across products, and/or store goods they purchase dur-

---

<sup>1</sup>Aguiar et al. (2013) show that about 30 percent of lost labor hours were reallocated toward non-market work, including shopping, during the Great Recession.

<sup>2</sup>Griffith et al. (2009) also document four channels of saving: Purchasing items when they are on sale, buying in bulk (at lower per-unit prices), buying generic brands, and shopping at outlets.

ing sale periods, and (2) “loyals,” who buy only one brand and do not time purchases to coincide with sales. In this setting, it is optimal for firms to maintain a combination of constant regular prices and frequent short-lived sales.

Kaplan et al. (2019) use the Nielsen scanner dataset and find that most variation in prices happens within stores rather than across them. They construct a model with two groups of consumers: (1) “busy,” who make all purchases in one store, and (2) “cool,” who shop at several stores. Under their assumptions, in equilibrium stores will charge different prices for the same goods. Intuitively, busy consumers will buy expensive and less-expensive goods in the same store, while consumers with more time for shopping will buy the cheaper goods at each store. Our paper is the first to document that individuals display different patterns of price-search across goods—for example, being attentive to prices, like the “cool shoppers,” when buying diapers and inattentive, like the “busy loyals,” when buying beer.

Macroeconomists have paid attention to rational inattention since Reis (2006) found that consumption patterns are consistent with the notion that consumers update information infrequently due to difficulty acquiring, absorbing, and processing information. A large literature (see Sims, 2010)—somewhat disconnected from the literature on bargain hunting—focuses on rational inattention, and its implications for monetary policy. Our work is related to this literature in that we posit that consumers pay less attention to goods of less importance to them.

Following Aguiar and Hurst (2007), we construct for each consumer a bargain-hunting index (BHI) which measures the price that he or she paid for a consumption bundle

relative to the cost of the same bundle based on average UPC prices in the same month and town. We refine the BHI to the category level, where a category (as defined by IRI) is a group of UPCs for similar goods, such as carbonated drinks. Studying behavior at the category level, rather than the UPC level, reduces noise significantly, because most consumers purchase only a small fraction of UPCs.<sup>3</sup>

Using the category BHI, we show that consumers who purchase “relatively more” units of goods in a category pay lower prices than the average consumer for goods within that category. An individual can consume relatively more in a category than other consumers in a given period  $t$ , the individual can consume relatively more in a category than his or her consumption in other categories, or the individual can consume relatively more at time  $t$  in a category than he or she consumes on average in that category. Using fixed effects, we show that our results are robust to various implicit definitions of “relatively more.” We also find, consistent with other studies, that prices paid differ substantially across consumers; in particular, retirees pay less and high-spending (“wealthy”) consumers pay more for identical baskets of goods. Quantities may be endogenous to prices so we estimate our relationships using instrumental-variable (IV-) regressions, although results from Ordinary Least Squares (OLS) are similar indicating that reverse causality is limited in our setup.

Building on earlier time-allocation models, such as those in Becker (1965), Benhabib et al. (1991), and Greenwood and Hercowitz (1991), we construct a simple static search model where consumers can trade off time versus prices good-by-good. The model predicts

---

<sup>3</sup>Figure 3 lists the included categories.

that consumers pay relatively less for goods of which they purchase relatively more as they optimally search relatively more for better prices of goods that they prefer. Also, consumers in our model may decide not to search at all for low prices, or only search for low prices of their preferred good. The model also predicts that individuals with high wages pay relatively more and that retired individuals on fixed income pay relatively less. We do not rely on time-use data in our empirical work and while our benchmark model is written in terms of a time constraint, our favored interpretation of the empirical results is that consumers obtain low prices by expending effort more broadly defined; for example, by paying attention to sales. In the appendix, we outline a model of store price-setting which is consistent with a model where consumers search only for low prices for their preferred goods.

The rest of the paper is organized as follows: Section 2 derives a simple model of time use. Section 3 describes the data. Section 4 presents our results and Section 6 concludes.

## **2 A Stylized Model of Consumer Search**

Income varies across consumers as do their preferences over two different goods and leisure. Each consumer faces a time constraint, where leisure is residual time after searching for good(s) and working. A consumer can spend time searching for lower prices; however, the consumer will rationally choose not to search, to search for better prices of just one good, or to search for better prices of both goods based on the utility of each choice. We formulate a simple model with differentiable search and utility functions, and with a

discrete choice on how many goods to search for. We solve the model for each of these choices and show graphically how a consumer may rationally decide between these discrete outcomes as a function of preferences and constraints.

Consider consumers, indexed by  $i$ , who search for best prices of both goods and derive utility as summarized by the objective function:

$$\max_{Q_i^1, Q_i^2, T_i^Y, T_i^1, T_i^2} \quad \alpha_i^1 \ln Q_i^1 + \alpha_i^2 \ln Q_i^2 + \mu \ln (T - T_i^Y - T_i^1 - T_i^2 - T_i^0), \quad (1)$$

$$\text{subject to:} \quad P_i^1 Q_i^1 + P_i^2 Q_i^2 = Y(T_i^Y), \quad (2)$$

where  $Q_i^1$  and  $Q_i^2$  are the purchased quantities of good 1 and good 2, and  $\alpha_i^1/\alpha_i^2$  is consumer  $i$ 's preference for good 1 relative to good 2, with  $\alpha_i^1 + \alpha_i^2 = 1$ .  $T_i^Y$  is the time devoted to income-generating activities (work).  $T_i^u$  ( $u = 1, 2$ ) is time spent searching for good (UPC)  $u$ , and  $T_i^0$  is a search fixed cost incurred whenever the consumer searches for any of the goods. "Search time" refers to time expended to obtain lower prices. Search time is used not just for price discovery: For example,  $T_i^0$  could be the time used reading advertisements at home, while  $T_i^u$  could be the time spent driving to the store where good  $u$  is purchased.  $T$  is the total time endowment, and  $T - T_i^Y - T_i^1 - T_i^2 - T_i^0$  is leisure. Solutions with negative time in any activity are not valid, but we do not explicitly write down Kuhn-Tucker conditions for ease of exposition.

Income is a linear function of work time, with wage rate  $W_i^1$ , and non-wage income,  $W_i^0$ :  $Y(T_i^Y) = W_i^0 + W_i^1 T_i^Y$ . Time spent searching results in lower prices according to the function  $P_i^u = P^h(T_i^u + \eta)^{-\beta}$ , with  $\beta > 0$ .  $\beta$  determines the efficiency of search: The

larger  $\beta$ , the more a unit of search time lowers the price paid.  $P^H = P^h \eta^{-\beta}$  is the highest price charged by stores when no search occurs. The marginal effect of an additional unit of search is  $\frac{dP_i^u}{dT_i^u} = -\beta P^h (T_i^u + \eta)^{-\beta-1} = -\beta P_i^u (T_i^u + \eta)^{-1}$ .

We discuss how the consumer splits his or her time next—detailed derivations, using a standard Lagrangian, can be found in an online appendix. Work time increases in total time available and in wages ( $W^1$ ), and it decreases in non-wage income ( $W^0$ ), price-search efficiency ( $\beta$ ), and leisure preference ( $\mu$ ).<sup>4</sup> The explicit solution for work time is

$$T_i^Y = \frac{T - T_i^0 + 2\eta - (\beta + \mu) \frac{W_i^0}{W_i^1}}{1 + \beta + \mu}. \quad (3)$$

Relative time spent searching for different goods is proportional to the utility weights, so agents rationally allocate more time to preferred goods. Search time increases with total time available, non-wage income and search efficiency, and decreases with wages and leisure preference. The solution is

$$T_i^u = \beta \alpha_i^u \frac{T - T_i^0 + 2\eta + \frac{W_i^0}{W_i^1}}{1 + \beta + \mu} - \eta. \quad (4)$$

Leisure time is

$$T - T_i^1 - T_i^2 - T_i^Y - T_i^0 = \frac{\mu}{1 + \beta + \mu} \left( T - T_i^0 + \frac{W_i^0}{W_i^1} + 2\eta \right).$$

For a fixed  $W_i^0$ , a larger wage,  $W_i^1$ , implies more work time, greater income, and less

---

<sup>4</sup>We assume that the parameters and  $W_i^0$  and  $W_i^1$  are such that  $T_i^Y$  is non-negative.

search time, so the “wealthy” (in terms of labor income) pay more. Work time decreases with the ratio  $W_i^0/W_i^1$ , and  $T_i^u$  increases, so retirees are predicted to pay lower prices. The higher the value of  $\eta$ , the less time is spent searching for low prices and the more time is spent on work and leisure. This reflects that the marginal return to searching is declining with  $\eta$ . For simplicity, assume  $W_i^0/W_i^1$  is such that work time is 0 for retirees. The unconstrained solution to the consumer problem may involve negative search time for the least preferred good, in which case the actual solution will be one with no search for that good. The appendix further describes consumers’ time allocation when they only search for best prices of one good or none.

## 2.1 Empirical Implications

The model implies an inverse relationship between the prices paid and the quantities consumed of a given good, because consumers vary their search intensity across goods in accordance with their relative preferences.

We highlight the importance of search fixed costs in Figure 1. The figure depicts (restricted) utility under three scenarios, each represented by a line in the figures: (1) the consumer does not search for better prices at all, (2) the consumer searches only for his or her preferred good (that with the highest  $\alpha^u$ ), and (3) the consumer spends time to obtain better prices for both goods. The consumer will evaluate the utility under these three scenarios and rationally choose the one delivering the highest utility. In the figure, we illustrate three cases: (a) the search fixed cost is zero; (b) the search fixed cost is positive and the same when searching for one or two goods ( $T^0 = 5$ ); and (c) the search fixed cost

is higher when searching for two goods ( $T^0 = 10$  vs.  $T^0 = 5$ ). We further vary the level of search efficiency ( $\beta = 0.25$  or  $\beta = 0.5$ ) and, in all figures, the relative preferences for good 1,  $\alpha^1$ .

Consider the case where the consumer cares for both goods equally ( $\alpha^1 = 0.5$ , in the middle of each figure), and the search fixed cost is low and/or the search efficiency is high. In these scenarios, the consumer is better off searching for lower prices of both goods. In contrast, when fixed costs are high and search efficiency is low, the consumer optimally decides not to search at all (see the left figure of Panel C). When the consumer has differential preferences for the two goods, he or she may optimally decide to spend time searching for low prices of just one good. In this illustration, this situation occurs for extreme relative preferences in Panel B, but also if the search for a second good entails an additional fixed cost, see Panel C.

In our empirical work, we consider consumers prices and quantities relative to other consumers. In order to depict the implications of the model for relative prices and quantities, we select  $N$  model-individuals with preferences  $\alpha_i^1$  uniformly distributed on the  $[0,1]$  interval. For each of these individuals, we calculate prices and quantities and calculate the average across the individuals. In Panel A of Figure 2, we plot, for each value of  $\alpha_i^1$ , the price paid for and quantity purchased of good 1 relative to the average by the agent with this preference value.

In Panel B, we plot the savings on good 1 of each consumer (indexed by the preference value) as a function of their relative quantity. Savings compares actual spending on good 1 to what this consumer would have paid, had he or she paid the average price. Specifically,

we construct a bargain-hunting index for good 1:

$$BH_i^1 = 1 - \frac{P_i^1 Q_i^1}{\bar{P}^1 Q_i^1},$$

where  $\bar{P}^1$  is the average price paid for good 1 across all agents, weighted by purchased quantities:  $\bar{P}^1 = \sum_{i=1}^N w_i^1 P_i^1$ , with  $w_i^1 = Q_i^1 / \sum_{i=1}^N Q_i^1$ . The denominator in the fraction is the money that would have been spent on good 1 if the consumer purchased the amount  $Q_i^1$  at average prices, while the numerator is the actual expense on good 1. One minus the ratio, therefore, shows the percent savings that the consumer obtains from searching for that good. We define a relative quantity index (purchased quantity relative to average quantity) as:

$$QI_i^1 = \frac{Q_i^1}{\frac{1}{N} \sum_{j=1}^N Q_j^1}.$$

Figure 2, Panel B, highlights the positive (negative) relationship between savings (relative prices paid) and relative quantities purchased. We focus on relative savings/prices and relative quantities across consumers going forward and generalizations of the bargain hunting and quantity indices will be prominent in our empirical regressions.

In summary, the model predicts an inverse relationship between prices and quantities when differential preferences for goods lead to differential amounts of time searching for low prices across goods. Other factors that affect the opportunity cost of shopping time (for example, wage levels) will also affect prices. In the empirical section, we interpret “search time” broadly because we do not have measures of literal search time. Consumers may expend effort by paying attention to prices (to determine, for example, whether

there are sales), consistent with the references to consumers' limited mental capacity in the rational inattention literature, or consumers may literally spend time driving to a larger selection of stores and/or comparing prices while at the store.

### **3 Data Description**

#### **3.1 The IRI Academic Dataset**

We use the IRI academic dataset which, as Bronnenberg et al. (2008) describe in detail, contains weekly transaction information on the purchases of groceries in 31 item categories. Our dataset spans 2001 through 2012 and includes information about purchases at the store level and at the individual level. At both levels, weekly total dollar and unit sales are collected for each UPC item. A UPC is encoded in a bar code used for scanning at the point of sale, and it contains information on very specific product attributes, such as volume, product type, brand name, package size, and even flavor or scent for some products. Products that are essentially the same but differ in size or packaging have different UPCs; for example, a bottle of Budweiser beer intended for single sale has a different UPC code than a physically identical bottle of Budweiser beer sold in a six-pack.

The store-level data contain weekly total-dollar and unit-sales information for each UPC from grocery stores and drug stores in 50 IRI markets (metropolitan areas). Most stores belong to large chains (masked), and each store has a unique identifier. The individual-level panel dataset provides price and quantity information for all transactions (where a "transaction" is a UPC-specific purchase) made by a consumer panel in two

small markets (cities): Eau Claire, Wisconsin, and Pittsfield, Massachusetts. The dataset includes some demographic information about the consumers, such as age, marital status, education, income, employment status, and family size. However, these variables are collected sporadically, reported only for discrete categories, and not consistently coded over time, so we include only a dummy for 65-plus years of age in our regressions.<sup>5</sup> Prices are linked to the store-panel data for purchases from the stores in the IRI store dataset. When IRI does not receive store data directly because the store is not in the set of stores followed, consumers record prices using an electronic wand.

The IRI dataset also includes a supplemental “trips file” that provides information on when (week) and where (store number) each panelist went shopping, as well as the amount of money spent while shopping. We calculate the total number of trips each panelist made to stores in a given period and the number of stores visited. We mainly use the individual-level transaction data, but we use price information from the store-level dataset to calculate average market prices by UPC.

We exclude the years 2001 and 2002 due to incomplete information and inconsistencies with later years, and we exclude “soup” purchases due to unrealistic price variation for this category—the exclusion of these years and this product category does not significantly affect our results, though. The store-level dataset is available for cities other than Eau Claire and Pittsfield, but we only make use of the data for these two markets because they can be linked to the consumer panel. For regressions on overall expenditure, we drop

---

<sup>5</sup>Our main results are not sensitive to inclusion of panelist fixed effects, which control for all non-time-varying consumer characteristics, so it is unlikely that including this information would alter our conclusions.

panelist  $\times$  month observations if the panelist's expenditure in the month is less than \$5. For regressions on category expenditure, we drop the panelist  $\times$  month  $\times$  category cell if the panelist's expenditure in the month is less than \$2 for that category.<sup>6</sup>

The appendix gives more details about the consumer panel, including the brackets in which income, age, and education are reported. Table A.1 displays summary statistics for the panelists in January, 2007. Average education is 13.8 years, and average age is 55 years. Individuals in our sample are between 21 and 70 years old. Average income is roughly \$52,500 with a standard deviation of \$36,600 (the standard deviation is likely lower than the actual standard deviation because income is reported in brackets). About a third of the sample is over 65. Average expenditure is about \$80 a month.

Compared with the Panel Study of Income Dynamics (PSID), a representative sample of the United States (for which we do not tabulate the numbers), the IRI panelists in 2007 are somewhat older (the average age of a PSID household head is 50), poorer (average income in the PSID is \$67,000), and similarly educated (the average number of years of education completed in the PSID is 13.1). In the PSID, the average food-at-home expenditure in 2007 is roughly \$4,400 a year. Using that number as an approximation of average food consumption for our sample, it implies that spending on categories in the panel IRI dataset constitutes 22 to 28 percent of food-at-home expenditure.<sup>7</sup>

---

<sup>6</sup>IRI includes only respondents who make at least one transaction in each of the 13 four-week periods in each year. (The documentation does not make this more precise.)

<sup>7</sup>The lower number does not adjust for income differences in the two samples, whereas the higher number does.

### 3.2 Data Construction

Similarly to Aguiar and Hurst (2007), we define the average price of a good  $u$  (UPC item), in a given market  $m$ , and month  $t$ , as a quantity-weighted average of the prices of all transactions  $k$  that involve that specific good. A transaction in our analysis is the purchase of a given good during a visit to a store, where one visit to a store usually comprises many transactions.<sup>8</sup> The average price of good  $u$  is

$$\bar{P}^{u,m,t} = \frac{\sum_{k \in u,m,t} Q_k^{u,m,t} P_k^{u,m,t}}{\sum_{k \in u,m,t} Q_k^{u,m,t}},$$

where  $Q_k$  is the quantity purchased in transaction  $k$  (involving  $u$ ), and  $P_k$  is the unit price. To compute this average price, we use the store-level dataset, which includes all transactions in all IRI stores in a given market. We refer to this price as the store-average price. The choice of a time-horizon of one month is arbitrary. Some goods can be easily stored for a month, allowing consumers to time their purchases independent of consumption, while others cannot be stored for more than a few days. Our results are qualitatively robust to whether we average prices over a month, a week, or a quarter. To keep it simple, we relate (average) prices to quantities purchased during the same period of time, a month. For shorter time-periods, there are many zero or near zero purchases which may result in less precise estimates.<sup>9</sup>

---

<sup>8</sup>For example, a purchase of three toothbrushes and a toothpaste tube is at least two transactions. It would be four transactions if the toothbrushes were all different brands, and even if the toothbrushes were of the same brand, there would be three transactions in total, if the consumer bought a single toothbrush and a two-pack.

<sup>9</sup>One may relate, say, paid prices to average weekly prices and aggregate purchased amounts to the quarterly level, as we did in an earlier version of this paper. The qualitative conclusions are similar, and

Following Aguiar and Hurst (2007), we define an overall bargain-hunting index for consumer  $i$  in period  $t$  ( $\text{BHI}_{i,t}$ ) as the amount a consumer saves for the products he or she buys relative to the cost of the exact same products at average prices in the same month and market. Specifically, the bargain-hunting index is computed as follows:

$$\text{BHI}_{i,t} = \left( 1 - \frac{\sum_{k=1}^{N_i^t} P_{i,k}^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}}{\sum_{k=1}^{N_i^t} \bar{P}^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}} \right) \times 100, \quad (5)$$

where  $i$  is a consumer who purchases products in market  $m$ ,  $k$  is a transaction of consumer  $i$ ,  $u(k)$  is the UPC of the transaction, and we aggregate expenditure to the monthly frequency  $t$  by summing over all  $N_i^t$  transactions of consumer  $i$  in month  $t$ . A consumer purchases many products and can purchase a particular product more than once a month, so the number of transactions is at least as large as the number of different goods purchased.

For each transaction, we use the exact price of the good in that transaction,  $P_{i,k}^{u(k),m,t}$ , to calculate actual expenditure. Given the consumer's consumption bundle, hypothetical expenditure is measured using the store-average price of the good in that transaction in the same market and month— $\bar{P}^{u(k),m,t}$  where  $u(k)$  denotes the UPC of the good purchased in transaction  $k$ . A higher BHI means saving more (paying less) relative to the store-average prices given the household's consumption bundle but, while this index encompasses all goods a household purchases, the prices are compared at the UPC level. Expenditure is

---

we prefer to limit confusion by aggregating quantities over a month to match the averaging frequency for prices. Results are similar when using quarterly frequencies to both aggregate quantities and to calculate average prices. See the online appendix.

aggregated over a month in order to avoid a large number of zero observations at higher frequencies (our results also hold for a quarterly frequency). We also calculate, for each consumer, the number of shopping trips as well as the number of different stores visited each month to study possible channels for bargain-hunting savings.

Table A.2 in the online appendix displays the mean and standard deviation of the BHI (along with summary statistics for other variables used in the regressions). The average BHI is 7.5 percent, which means that panelists save 7.5 percent on average by finding better-than-average prices. The average price for each UPC is calculated outside the panelist sample and includes transactions by all shoppers in these markets; a positive average could reflect that panelists in our sample are, on average, older than the typical population.<sup>10</sup> It is also possible that stores outside the IRI store sample are on average cheaper. Going forward, we demean the BHI to 0 each period.<sup>11</sup>

Our main focus is on selective bargain hunting (that is, whether consumers devote relatively more time to search for lower prices for goods they prefer). Our model predicts that consumers pay relatively less for preferred goods. Testing the relation between prices and goods at the UPC level by regressing prices on quantities is, however, complicated by potential endogeneity of quantities, and because almost all households purchase just a few different UPC-level goods in a given period. We therefore use a modified version of the bargain-hunting index defined at the category level (for each consumer and time period), and a quantity index by category. The quantify index measures whether a consumer buys

---

<sup>10</sup>The un-weighted average over a set of consumers can also deviate from the quantity weighted average for the same consumers when quantities vary across consumers.

<sup>11</sup>Demeaning is not strictly necessary in our regression analysis, because we include period (year  $\times$  month) fixed effects.

relatively more or less of that category. Categories are defined by IRI as groups of similar goods, a full list is provided in Figure 3, but two examples are (non-alcoholic) carbonated beverages and beer.<sup>12</sup>

Some consumers may have very strong preferences for a particular type of carbonated beverage at the UPC level and only purchase, for example, Diet Coke cans in “fridge packs” of twelve. Other consumers may like Budweiser beer, but not be particular about package size or type (e.g., cans versus bottles). Such consumers may purchase whichever Budweiser is cheapest. Other consumers may like beer, not caring too much about brand nor packaging. For the latter two groups of consumers, quantities at the UPC level will be highly endogenous to prices. We therefore conduct our analysis at the category level and construct quantity indices at that level, which are less likely to be endogenous. For consumers that only consume one UPC in each category, the analysis will follow that of our stylized model as the category corresponds to the UPC, while for other consumers, the logic of the model will carry through.<sup>13</sup> Individuals may not substitute between beer and razors, but consumers may time their purchases leading to endogeneity problems at the category level, and our empirical analysis will utilize Instrumental Variable (IV) techniques. However, we find that our results are robust to the use of IV versus OLS.

Let  $c$  denote a category. A category-level BHI for a given consumer  $i$  in period  $t$ ,  $BHI_{i,t}^c$ , is computed similarly to the overall BHI, except that only transactions involving

---

<sup>12</sup>We could construct groups of UPCs ourselves, but there is no obvious way of doing this and an arbitrary choice of categories would open up the scope for data mining. Thus, we utilize the categories defined by IRI.

<sup>13</sup>One could write down a version of the model where goods within categories are perfect or imperfect substitutes. This would complicate the presentation for little purpose as we do not estimate elasticities of substitution, so we do not go that route.

products in a given category are added up:

$$\text{BHI}_{i,t}^c = \left( 1 - \frac{\sum_{k,u(k) \in c} P_{i,k}^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}}{\sum_{k,u(k) \in c} \bar{P}^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}} \right) \times 100. \quad (6)$$

Figure 3, Panel A, presents a box plot of the BHI by category, illustrating the range of prices paid, and thus consumers' potential for saving by searching, which varies by category.<sup>14</sup> In the graph, IRI's categories are ordered by the interquartile range of the category-specific BHIs. For example, the interquartile range for beer is 1/11th of that for laundry detergent (2.63 percent versus 28.92 percent). This significant difference is likely due to very disparate pricing strategies employed by retailers and/or producers of the two products; nevertheless, there is price variation for identical UPCs within all product categories, implying potential gains from price-search.

A category-level quantity index for a consumer  $i$  in period  $t$ ,  $\text{QI}_{i,t}^c$ , is computed as the value of his or her transactions in a given category relative to the average value across consumers of transactions in the same category. Both values are computed at average prices so that the resulting ratio reflects differences in quantities and not prices. Specifically:

$$\text{QI}_{i,t}^c = \frac{\sum_{k,u(k) \in c} \bar{P}^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}}{(\sum_{j \in J_t^m} \sum_{k,u(k) \in c} \bar{P}^{u(k),m,t} \times Q_{j,k}^{u(k),m,t}) / J_t^m}, \quad (7)$$

where  $J_t^m$  is the number of consumers in the panel in market  $m$  in period  $t$ ,  $k$  is an index

---

<sup>14</sup>The category names are intuitive, except maybe the category "blades," which is mainly made up of cartridges for shavers.

for transactions, and  $u(k)$  is the UPC of transaction  $k$ . The fixed price weights reflect differences in quantity and quality (broadly defined), so purchases of larger amounts of more expensive UPCs have greater weights than purchases of larger amounts of less expensive UPCs. This calculation of quantities purchased aligns with the model, because consumers have a stronger incentive to search for lower prices of goods that are, on average, more expensive.<sup>15</sup> Two individuals with similar values for the QI may differ in that one systematically buys fewer, but more expensive UPCs, than the other. Panel B of Figure 3 illustrates that there is substantial variation in the quantity index (winsorized at the top and bottom 1 percent) by category (ordered by interquartile range). By construction, the mean for the quantity index for each category is (roughly) 1.

The QI gives higher weight to more expensive UPCs and we do not compare prices across UPCs when computing the BHI. Nevertheless, we verified that individuals with lower QIs are not systematically buying more expensive UPCs on average. To do this, we calculate the average price of the UPCs purchased by individuals in a given category and correlate this price with the QI. The raw correlation is relatively low, 0.15, and positive. We also construct an equal-weight QI based on the total count of UPCs purchased in a given month by category (relative to the average count of UPCs in that category in each market and month).<sup>16</sup> The correlation of this equal-weight QI with the average price

---

<sup>15</sup>Neiman and Vavra (2019) show that consumers increasingly concentrate their consumption on individual-specific goods. Our quantity index is robust to such changes as it is invariant to concentration within the category. Neiman and Vavra (2019) construct a model where tastes for individual goods vary across consumers while there is a utility cost of consuming a large number of goods. Conceivably, such costs might be rationalized by search costs.

<sup>16</sup>In general, we are not able to transform purchases into a common unit of measurement, say, gallons. This is easily done for milk, but no so easily for other categories where the UPCs are more heterogenous.

of the UPCs in a given category is  $-0.02$ . In addition, the correlation of the QI and the equal-weight QI is high,  $0.7$ , and all of our results go through if we use the equal-weight QI in the regressions, so substitution between higher- or lower-price UPCs within categories is not likely to affect our results.

To explore whether consumers save from visiting certain stores or from timing of purchases, we compute for each consumer an alternative “store BHI,”  $\text{BHI}_{i,t}^{c,s}$ , that computes the value of consumer  $i$ ’s basket using the average price,  $\bar{P}_s^{u,m,t}$ , of each UPC in a given month *in the store*,  $s$ , where the item was purchased. To compute this average price, we use the store-level dataset.<sup>17</sup> The exact expression for the store (category-level) bargain-hunting index is

$$\text{BHI}_{i,t}^{c,s} = \left( 1 - \frac{\sum_{k,u(k) \in c} \bar{P}_s^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}}{\sum_{k,u(k) \in c} \bar{P}^{u(k),m,t} \times Q_{i,k}^{u(k),m,t}} \right) \times 100. \quad (8)$$

The store bargain-hunting index replaces the numerator (the amount paid for the purchased basket) in the bargain-hunting index with the counterfactual amount that the consumer would have paid for the purchased basket, had he or she paid the average price (in that month) in the store in which each good was purchased. This index is informative about whether the consumer obtains lower prices by shopping in stores where the desired goods are relatively cheap with the discrepancy to the regular bargain-hunting index explained by timing of purchases within stores. If the BHI for a consumer is lower than

---

<sup>17</sup>This exercise is performed using data from 2003 through 2007, because store identifiers in the store-level dataset are not fully consistent with identifiers in the panelist dataset after 2007. Also, goods purchased at stores outside the IRI sample are not included in the index.

the store BHI, the consumer has, on average, purchased goods at times of the month when the prices of the goods were lower than store-specific monthly average prices.

Figure A.3 in the appendix compares the original BHI to the store BHI using histograms. The correlation between the two indices is 0.52, and the histograms in Panel A suggest gains from both store selection and the timing of purchases. In Panel B, we display the histogram of savings by individuals by time of purchase; that is, we show the percentage saved by paying the actual price for each transaction rather than paying the store-monthly average for the relevant UPC. The distribution in Panel B includes many more observations of positive savings than of negative savings, indicating overall gains from the timing of shopping.

## 4 Empirical Results

In this section, we first report regressions similar to those in the bargain-hunting literature. By confirming previous results, we establish that our data does not deliver deviating results along the dimensions where we can compare to previous work. Second, we verify that consumers on average pay less for goods in preferred categories and, third, we show that consumers who purchase relatively larger quantities pay less in almost all categories.

In Table 1, we first report results from regressions of the form

$$\text{BHI}_{i,t} = \mu_i + \gamma_{m,t} + X_{i,t} \alpha + \epsilon_{i,t},$$

where  $\text{BHI}_{i,t}$  is the bargain-hunting index for individual  $i$  in month  $t$ ,  $\mu_i$  is an individual

fixed effect included in only some of the specifications,  $\gamma_{m,t}$  is a market  $\times$  month fixed effect, and  $X$  is a vector of regressors: A dummy for age 65 and older, and the logarithm of expenditure (our proxy for labor income). In some specifications, we also include the total number of shopping trips and the number of different stores visited in a given month. Aguiar and Hurst (2007) use a similar specification, although they do not include the number of stores visited.

The left panel of Table 1 shows results for regressions without individual fixed effects. The results in column (1), when only expenditure and age (besides market  $\times$  month fixed effects) are included, confirm previous results that higher-spending consumers pay relatively more (with an elasticity of  $-0.62$ ), and that consumers 65 and older pay relatively less (with average savings of 0.71 percent). This is consistent with the model's interpretation that high-wage workers elect to search less because of their higher value of time, while older individuals search more because of their lower value of time.

In columns (2)–(4), we include the number of different stores visited in a month and the number of shopping trips as direct measures of search effort. These variables are not necessarily exogenous—one might imagine some stores having frequent sales, which makes consumers take more trips to the store and obtain lower prices—but we include them because they are informative of the mechanism by which consumers obtain bargains. The number of stores visited in a given month predicts lower prices paid robustly and with high statistical significance. The economic interpretation of the coefficient to this variable in column (2) is that consumers who visit one more store each month pay 0.77 percent less for their consumption basket than they would have paid at average prices. The inclusion

of the number of stores visited increases the R-square from 0.01 to 0.04, so this variable has much greater explanatory power than do age and expenditure (although this likely reflects that the age dummy is somewhat imprecisely correlated with retirement, and it may be that retired people save more by visiting more stores). Including the number of shopping trips, in column (3), while omitting the number of stores visited, gives a highly significant coefficient for trips of 0.19 with an R-square of 0.02. However, including the number of shopping trips together with the number of stores visited, in column (4), lowers the coefficient to the number of shopping trips significantly, while the coefficient to the number of stores visited is quite similar across columns. Clearly, it is the number of stores visited, rather than the number of trips, that is associated with lower prices although, according to column (4), one more trip (controlling for number of stores visited) still lowers the average price paid by 0.05 percent.<sup>18</sup>

In the right panel of Table 1, we include individual fixed effects. The R-square jumps to 0.25, so it appears that some consumers are consistently “shoppers,” while others are “loyals” (in the parlance of Chevalier and Kashyap, 2019).<sup>19</sup> Expenditure and age are still significant. The coefficient on age is smaller as it is now identified only from consumers who turn 65 during the sample period. The number of stores visited remains significant, but with individual fixed effects, the coefficient drops to 0.22 in column (8). This indicates that some consumers consistently shop at many stores and these consumers may be more

---

<sup>18</sup>This is consistent with the findings of Kaplan et al. (2019) that some stores are cheaper for some goods but not for others.

<sup>19</sup>Our results are not informative about whether some consumers are inherently looking for deals or whether they, for instance, live close to an inexpensive store, or they are impacted by other unmeasured features, so we interpret those terms broadly.

informed, obtaining bigger savings from multi-store shopping. The number of trips and the number of stores visited do not add much to the explanatory power of the regressors when individual fixed effects are included.

The main innovation of this paper is that it examines the relationship between quantities and prices by category—which means that we study the basket of goods purchased within each category, although we base the comparison on UPC-level prices paid by the consumer versus average UPC-level prices across consumers. The bargain-hunting index measures the savings obtained by consumer  $i$  by paying below average prices for (UPC) good  $u$  in category  $c$ . The index is unit free and measures savings from search in percentages that are comparable across categories and households.<sup>20</sup>

The data form a panel indexed by individual  $\times$  category  $\times$  time and we estimate the regression

$$\text{BHI}_{i,t}^c = \nu_c + \gamma_{m,t} + \beta \text{QI}_{i,t}^c + X_{i,t} \alpha + \epsilon_{i,t},$$

allowing for category,  $\nu_c$ , and market  $\times$  time fixed effects,  $\gamma_{m,t}$ . The panel is highly unbalanced as a large number of consumers do not make purchases in each category each month and such cells are excluded from the sample. The focus of our paper is on the coefficient  $\beta$ ; in particular, our hypothesis is that  $\beta$  is positive, implying that consumers save relatively more in categories of which they purchase relatively more.

Quantities may be endogenous to prices; for example, if a consumer were indifferent between Coke and Pepsi and encountered a low price of a 12-pack of Coke, he or she

---

<sup>20</sup>The alternative of using budget shares is not feasible with our data because many other goods and services purchased are not observed.

would purchase that UPC, rather than Pepsi. This is fully consistent with our model, as consumers need to pay attention to prices in order to substitute between goods this way. Nonetheless, we provide a likely more convincing test of the model by showing that consumers find lower prices for UPCs in categories that they consume more of on average over the sample period. We therefore estimate the relationship using IV-regressions.

Regressions on category-level BHIs, rather than UPC-level BHIs, can in itself be seen as reduced-form IV-regression using the category dummies as instruments. For that instrument to be valid, the identifying assumption is that a consumer does not switch between categories; for example, buying beer instead of Coke when coming across a sale of Heineken. We find that assumption reasonable, although it could still be that someone who likes beer will stock up on enough Heineken to increase the value of the beer-category quantity index for the month when encountering (relatively) inexpensive Heineken. We, therefore, use the consumer's average category consumption as the instrument. Actually, in order to further hedge against reverse causality, our instrument is the average category-specific quantity index for each consumer over the sample period excluding the month in which the dependent variable is measured. The precise definition of this instrument is  $\overline{\text{QI}}_{i,t-}^c = \frac{1}{T_i-1} \sum_{s=1, s \neq t}^{T_i} \text{QI}_{i,s}^c$  where  $T_i$  is the number of months that consumer  $i$  is in the sample. The instrument is time-varying, but for simplicity we refer to it as the average category-specific quantity index.

The  $\beta$ -coefficient on the quantity index is (ignoring the controls, which are not of importance for this issue) identified from deviations from the fixed effects included. Including a market  $\times$  time fixed effect implies that differences between markets at any time do not

contribute to identification and, because of the category fixed effect, neither do permanent differences across categories. This leaves variation across consumers, across categories for each consumer, and across time within each consumer's category. We display IV results without and with individual fixed effects, which absorb (average) differences between individuals, and we display OLS results from estimations with category  $\times$  individual fixed effects that imply identification from time varying consumption within each consumer's category spending.<sup>21</sup>

The results presented in Table 2 show that consumers indeed pay less for goods in categories of which they purchase relatively larger quantities, consistent with rational allocation of time and effort across good categories. The coefficient to the quantity index is robustly estimated with a value of 2.39 in column (1), with similar magnitudes in columns (3)–(6). The coefficient is somewhat smaller at 1.45 in column (2), which includes individual fixed effects. In this case, the coefficient is identified from deviations from each consumers average consumption. All t-statistics are magnitudes larger than the 1 percent critical value, and the F-statistics for significance of the instruments are way beyond any reasonable critical value implying that the instruments are not weak.

The coefficient to age is significant and similar to what was found in Table 1, while the coefficient to log-expenditure is highly significant with an elasticity between  $-1.06$  and  $-1.47$  in the IV regressions. These coefficients are numerically slightly larger than those found in Table 1, which suggests that the former estimates may suffer slightly from

---

<sup>21</sup>Our favored instrument is highly correlated with the category  $\times$  individual fixed effects. Results are not very different if we use IV regressions with these additional fixed effects, but we prefer to display the OLS coefficient in this case.

left-out variable bias due to expenditure levels being correlated with the left-out category quantity index.<sup>22</sup>

A reduced-form OLS-regression with the average category-specific quantity index standing in for the current quantity index, see column (4), delivers a slightly smaller coefficient that is not directly comparable with the coefficients in the other columns, but the coefficient is highly significant and the interpretation remains that consumers find better prices for goods within desired categories. The standard OLS-regression coefficient in column (5) is of similar magnitude to the IV-coefficients in columns (1) and (3), indicating that reverse causality is likely not a problem for estimating the coefficient to the quantity index. Finally, in column (6), we show results for an OLS-regression including individual  $\times$  category fixed effects. In this regression, the coefficients are identified from temporal deviations from each consumer's quantity average; however, the estimated coefficient is insignificantly changed from column (5).

Table 3 explores a few issues. In column (1), the number of trips and the number of stores visited are included. The results for these variables are similar to those of Table 1. For example, the coefficient to number of stores visited in column (1) takes a value of 0.56 compared with a value of 0.68 in column (4) of Table 1. The similarity is to be expected unless the number of stores visited was correlated with the category dummies (which are not defined in the regressions using the overall bargain-hunting index). Even controlling for trips and the number of stores visited, the coefficient to the quantity index is very

---

<sup>22</sup>The regressions in Table 1 involve data aggregated over the categories. Left-out variable bias in the more disaggregated regressions may well translate to the aggregate level, but we will not pursue this issue in detail.

similar to what we found in previous regressions without these controls.

Search models typically imply that consumers search more when the dispersion in prices is higher. We check if results are robust to controlling for price dispersion within each category (by market and month). We compute the coefficient of variation (CV) of the prices of each UPC in a given market and month and average the UPC-level CVs across the UPCs in each category, creating a category-level CV.<sup>23</sup> We add this category-level CV as a control in our regressions as well as its interaction with the QI—the category-level CVs are standardized to have mean 0 and a standard deviation of 1 across categories for easier interpretation. Column (2) of Table 3 reveals that the quantify index has a stronger impact on prices in categories with higher price variation, consistent with standard search models. Consumers save more in categories of which they purchase relatively larger quantities, and more so if price dispersion is larger in the category.

In column (4), we show results using the store bargain-hunting index, which is calculated under the counterfactual assumption that the consumer paid the average price in the store for each good purchased in that store. (Column (3) repeats the regressions of Column (1) on the shorter sample and verifies that the differences in results are not simply reflecting the change in sample.) We run the regression

$$\text{BHI}_{i,t}^{c,s} = \nu_c + \gamma_{m,t} + \beta \text{QI}_{i,t}^c + X_{i,t} \alpha + \epsilon_{i,t},$$

which is similar to the regression of column (1), except for the left-hand side being the

---

<sup>23</sup>A simple average of the UPC-level CVs in a given category and averages that give more weight to higher-price UPCs or UPCs that are purchased more frequently deliver similar results.

store bargain-hunting index by category. For data reasons, we are only able to calculate the store BHI pre-2008, so the regression of the first column is repeated in column (2) for this truncated sample in order to make sure that estimated differences are due to the change in the index rather than the change in the sample. (The coefficients to the bargain-hunting index are virtually identical in the first two columns.) The results in column (4), obtained for the store BHI, reflect the prices the consumer *would have paid* if he or she had paid the average store price for the items purchased in the given month, rather than the actual price paid. Any difference to the results using the original BHI index is due to gains from the timing of purchases. The estimated coefficients for the store BHI are significantly smaller than those of the previous columns in both economic terms and in terms of statistical significance (the difference between the coefficients is in the order of a hundred times larger than the standard errors of the estimates). Specifically, the coefficient to the quantity index drops from 2.28 to 0.52, although both coefficients are extremely significant in a statistical sense. Our interpretation is that a large fraction of the savings obtained in favorite categories results from choosing the time to shop in a given store, rather than from choosing to shop in stores with consistently lower prices.<sup>24</sup>

The regression results reported in Tables 2 and 3 are pooled across categories, but pooling may mask differences across categories. As shown in Figure 3, Panel A, the BHI is significantly more compressed for some categories than for other categories, with almost no variation in prices paid for identical beer UPCs (on average) and little variation

---

<sup>24</sup>The results regarding the relative importance of timing is sensitive to the length of the period used. If, for example, store prices were averaged over only a week, the timing would be less important than the choice of store visited because there would be less scope for within-store price variation. The interpretation of this pattern is that some consumers are able to time their purchases over intervals longer than a week.

for cigarettes (followed by milk and sugar substitutes). Laundry detergent displays the largest variation, followed by hot dogs and frozen pizza. This is not a simple reflection of relative quantities consumed. For the quantity index, razors and ketchup/mustard show the least variation across consumers, and carbonated beverages and cigarettes the most (see Figure 3, Panel B). To test whether our results are robust across categories, we estimate the regression separately for each category. Table A.3 in the online appendix shows the regressions category-by-category and our main qualitative result is remarkably robust—the coefficient to the quantity index is positive and highly significant in almost all categories.

## 5 Discussion

As Aguiar and Hurst (2007), we compare prices at the UPC level, while we ignore savings from switching between brands or even package sizes. Because consumers may save by changing to less expensive brands or by buying larger packages of the same good, our findings provide a lower bound on the savings obtained by consumers. However, switching to different brands or package sizes brings a potential loss in utility that can be evaluated only by using functional forms, which we avoid in this paper.

If a consumer changed consumption within a category to more expensive brands, this would appear as an increase in the quantity index. The bargain-hunting index compares the price paid for a brand to the average price of that particular brand (or rather UPC, which is even more specific), so our model, as implemented, predicts that a consumer

that switches to a more expensive brand will (on average) search more for a relatively low price of that brand. To the extent that this does not happen, it will play the role of measurement error and bias our coefficients towards zero.

Our results are not only capturing the effects from permanent preferences across goods, as regressions with individual  $\times$  category fixed effects, which absorb permanent differences across consumers and across categories for each consumer, deliver similar results. The interpretation is therefore that if a consumer bought a lot of, say, beer either on average or in a month, for example due to a celebration, the consumer would search for lower prices for his or her preferred beer.

Our consumer model assumes different a priori preferences for goods, but one can imagine a case where a consumer has no preference for bananas versus apples but buys the fruit that happens to be relatively cheaper when he or she visits the store, which would result in an inverse relationship between good prices and quantities. (This, of course, is also rational behavior on behalf of the consumer, but the store-pricing implications would differ.) We use IV-regressions for this very reason. We find very similar IV- and OLS-estimates, which indicates that causality goes from preferences to prices after aggregating to the category level. In addition, we would expect random (for the consumer) sales to average out over a longer time period, and we, as a second indicator that most causality goes from preferences to prices, repeat the regressions in Table 3 for quarterly frequencies. The results, show in Table A.4, are very similar to those obtained at the monthly frequency, which supports the causal interpretation of the results.

## 6 Conclusion

We find that, consistent with a model of rational price search, consumers pay lower-than-average prices for goods of which they consume relatively more and they pay higher-than-average prices for goods of which they consume relatively less. The empirical results provide robust support for the notion that consumers rationally search for better prices when it has a higher return.

Our results are consistent with those from models of “shoppers” versus “loyals,” or “cool” versus “busy” consumers, in that we document significant variation in the prices consumers pay. In Table 4, we illustrate the magnitudes of the savings. The top quarter of consumers in the BHI distribution (the cool shoppers) pay, on average, 11.75 percent less than the average consumer, whereas consumers in the bottom quarter (the busy loyals) pay 10.21 percent more for the exact same goods. The main innovation of our work is that we depart from the assumption that some consumers pay low (high) prices across the board. For each consumer, we rank his or her purchased categories according to the quantity index and divide the goods into top-half and bottom-half categories (for this exercise, we include only consumers who purchase at least two categories).<sup>25</sup> We then compute separate BHIs for top-half and bottom-half categories (in terms of the quantity index). On average, consumers save 0.74 percent on the goods of which they buy more (relative to other consumers) and pay 2.28 percent more for the goods of which they buy less. Most savings accrue to consumers who find low prices across the board, although these consumers clearly obtain higher savings for their most preferred goods. Consumers

---

<sup>25</sup>If the number of categories is not even, the top group has one more category.

who on average pay the highest prices show no tendency for saving more on preferred goods, though.

Overall, there is substantial heterogeneity across consumers, as previously documented. Our contribution is to study how rational consumers shop across goods, and show that consumers conduct more bargain hunting for their most desired goods. An additional implication of our model is that improvements in search efficiency, while lowering prices on average, may not result in lower price dispersion across consumers. Given that consumer preferences for a particular good differ, search efforts will still vary across categories and price dispersion may not decrease. (We illustrate this numerically in Figure A.4.)

## References

- Aguiar, M. and E. Hurst (2007, December). Life-Cycle Prices and Production. *American Economic Review* 97(5), 1533–1559.
- Aguiar, M., E. Hurst, and L. Karabarbounis (2013, August). Time Use during the Great Recession. *American Economic Review* 103(5), 1664–1696.
- Becker, G. (1965). A Theory of the Allocation of Time. *Economic Journal* 75(299), 493–508.
- Benhabib, J., R. Rogerson, and R. Wright (1991, December). Homework in Macroeconomics: Household Production and Aggregate Fluctuations. *Journal of Political Economy* 99(6), 1166–1187.
- Bronnenberg, B. J., M. W. Kruger, and C. F. Mela (2008, 07-08). The IRI Marketing Data Set. *Marketing Science* 27(4), 745–748.

- Chevalier, J. A. and A. K. Kashyap (2019). Best Prices: Price Discrimination and Consumer Substitution. *American Economic Journal: Economic Policy* 11(1), 126–159.
- Coibion, O., Y. Gorodnichenko, and G. H. Hong (2015, March). The Cyclicity of Sales, Regular and Effective Prices: Business Cycle and Policy Implications. *American Economic Review* 105(3), 993–1029.
- Greenwood, J. and Z. Hercowitz (1991, December). The Allocation of Capital and Time over the Business Cycle. *Journal of Political Economy* 99(6), 1188–1214.
- Griffith, R., E. Leibtag, A. Leicester, and A. Nevo (2009, Spring). Consumer Shopping Behavior: How Much Do Consumers Save? *Journal of Economic Perspectives* 23(2), 99–120.
- Kaplan, G. and G. Menzio (2013, November). The Research Agenda: Greg Kaplan and Guido Menzio on the Macroeconomics of Bargain Hunting. *Economic Dynamics Newsletter* 14(2).
- Kaplan, G., G. Menzio, L. Rudanko, and N. Trachter (2019). Relative Price Dispersion: Evidence and Theory. *American Economic Journal: Microeconomics* 11(3), 68–124.
- Neiman, B. and J. S. Vavra (2019, August). The Rise of Niche Consumption. NBER Working Papers 26134, National Bureau of Economic Research, Inc.
- Nevo, A. and A. Wong (2019). The Elasticity of Substitution Between Time and Market Goods: Evidence from the Great Recession. *International Economic Review* 60(1), 25–51.
- Reis, R. (2006, November). Inattentive Consumers. *Journal of Monetary Economics* 53(8), 1761–1800.

- Salop, S. and J. Stiglitz (1977). Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion. *Review of Economic Studies* 44(3), 493–510.
- Sims, C. A. (2010). Rational Inattention and Monetary Economics. In B. M. Friedman and M. Woodford (Eds.), *Handbook of Monetary Economics Vol. 3*. Amsterdam, The Netherlands: Elsevier.
- Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy* 69(3), 213–225.
- Stroebel, J. and J. Vavra (2014, November). House Prices, Local Demand, and Retail Prices. NBER Working Paper No. 20710, National Bureau of Economic Research, Inc, Cambridge, MA.

Table 1. The Bargain-Hunting Index for Overall Expenditure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pooled OLS				Panelist Fixed Effects			
Log. Expenditure	-0.62*** (0.05)	-0.81*** (0.05)	-1.08*** (0.05)	-0.91*** (0.05)	-0.55*** (0.03)	-0.66*** (0.03)	-0.80*** (0.03)	-0.78*** (0.03)
Old (65+)	0.71*** (0.12)	0.40*** (0.11)	0.34*** (0.11)	0.35*** (0.11)	0.36** (0.17)	0.30* (0.17)	0.32* (0.17)	0.29* (0.17)
# stores visited (monthly)		0.77*** (0.03)		0.68*** (0.03)		0.32*** (0.02)		0.22*** (0.02)
# trips (monthly)			0.19*** (0.01)	0.05*** (0.01)			0.10*** (0.01)	0.06*** (0.01)
Month-Year $\times$ Market FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	551438	551438	551438	551438	551438	551438	551438	551438
Adj. R-squared.	0.01	0.04	0.02	0.04	0.25	0.25	0.25	0.25

Notes: Regression:  $BHI_{i,t} = \mu_i + \gamma_{m,t} \alpha + \epsilon_{i,t}$ , where  $BHI_{i,t}$  is the bargain-hunting index for individual  $i$  in month  $t$ ,  $\mu_i$  is an individual fixed effect (FE) ( $\mu_i = \mu$  in the first four columns),  $\gamma_{m,t}$  is a market  $\times$  month FE, and  $X$  is a vector of regressors: A dummy for age 65 and older, the logarithm of total expenditure, the number of different stores visited monthly, and the total number of shopping trips in the month. Standard errors clustered by individual. \*\*\* (\*\*\*) [\*] significant at the 1 (5) [10] percent level.

**Table 2. Rational Inattention. Pooled Regressions**

	(1)	(2)	(3)	(4)	(5)
Instrument:	IV avg. QI	IV avg. QI	OLS avg. QI	OLS current QI	OLS current QI
Quantity Index	2.39*** (0.07)	1.45*** (0.04)	1.99*** (0.06)	2.32*** (0.02)	2.28*** (0.02)
Log. Expenditure	-1.47*** (0.04)	-1.06*** (0.03)	-0.76*** (0.05)	-1.43*** (0.05)	-1.55*** (0.02)
Old (65+)	0.42*** (0.11)	0.24* (0.14)	0.62*** (0.11)	0.42*** (0.11)	0.23* (0.13)
Month-Year $\times$ Market FE	Yes	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	Yes	Yes	No
Individual FE	No	Yes	No	No	No
Category $\times$ Individual FE	No	No	No	No	Yes
Observations	3768440	3768440	3768440	3793575	3768440
Adj. R-squared.	0.02	0.02	0.01	0.02	0.15
F-test excl. inst.	1331271	1261199			

*Notes:* Regression:  $BHI_{i,t}^c = \nu_c + \gamma_{m,t} + \beta QI_{i,t}^c + X_{i,t} \alpha + \epsilon_{it}$ , where  $BHI_{i,t}^c$  is the category-specific bargain-hunting index for individual  $i$  in month  $t$ , and  $\nu_c$  denotes category fixed effects.  $\gamma_{m,t}$  is a market  $\times$  month fixed effect,  $X$  is a vector of regressors, and  $QI_{i,t}^c$  is the quantity index described in equation (7). The quantity index, which measures whether a consumer purchases more or less of a category than the average consumer, is standardized (mean 0, sd 1) for easier interpretation. The IV-regressions use as an instrument the average category-specific quantity index defined for observation  $t$  as  $\bar{QI}_{i,t-}^c = \frac{1}{T_i-1} \sum_{s=1, s \neq t}^{T_i} QI_{i,t}^c$ . Standard errors clustered by individual. \*\*\* (\*\*) [\*] significant at the 1 (5) [10] percent level.

**Table 3. Rational Inattention. IV. Trips and Stores Controls.  
Store Index.**

	(1)	(2)	(3)	(4)
	All Years	All Years	Pre-2008	Pre-2008
	BHI	BHI	BHI	Store BHI
Quantity Index	2.28*** (0.06)	2.30*** (0.06)	2.36*** (0.06)	0.52*** (0.03)
Log. Expenditure	-1.66*** (0.05)	-1.67*** (0.05)	-1.77*** (0.05)	-0.60*** (0.02)
Old (65+)	0.08 (0.10)	0.09 (0.10)	-0.17 (0.10)	-0.10** (0.05)
# stores visited (monthly)	0.56*** (0.03)	0.56*** (0.03)	0.76*** (0.03)	0.33*** (0.01)
# trips (monthly)	0.05*** (0.01)	0.05*** (0.01)	0.05*** (0.01)	0.01** (0.00)
QI × Prices-CV		0.50*** (0.04)		
Prices-CV		-0.00 (0.02)		
Month-Year × Market FE	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	Yes	Yes
Observations	3729149	3729149	2133968	2133968
Adj. R-squared	0.03	0.03	0.04	0.01
F-test excl. inst.	1315054	644147	799500	799500

*Notes:* Regression:  $BHI_{i,t}^c = \nu_c + \gamma_{m,t} + \beta QI_{i,t}^c + X_{i,t} \alpha + \epsilon_{it}$ , where  $BHI_{i,t}^c$  is the category-specific bargain-hunting index for individual  $i$  in month  $t$ , and  $\nu_c$  denote category fixed effects.  $\gamma_{m,t}$  is a market × month FE,  $X$  is a vector of regressors, and  $QI_{i,t}^c$  is the quantity index described in equation (7). The quantity index, which measures whether a consumer purchases more or less of a category than the average consumer, is standardized (mean 0, sd 1) for easier interpretation. Prices-CV denotes the average of the coefficients of variation of all UPC-level prices in a given category by market and month. In column (3),  $BHI_{i,t}^c$  is replaced by a category-specific store BHI,  $BHI_{i,t}^{c,s}$ . All regressions are estimated by IV, using the average category-specific quantity index as an instrument, defined more precisely in the notes to Table 2. Standard errors clustered by individual. \*\*\* (\*\*) [\*] significant at the 1 (5) [10] percent level.

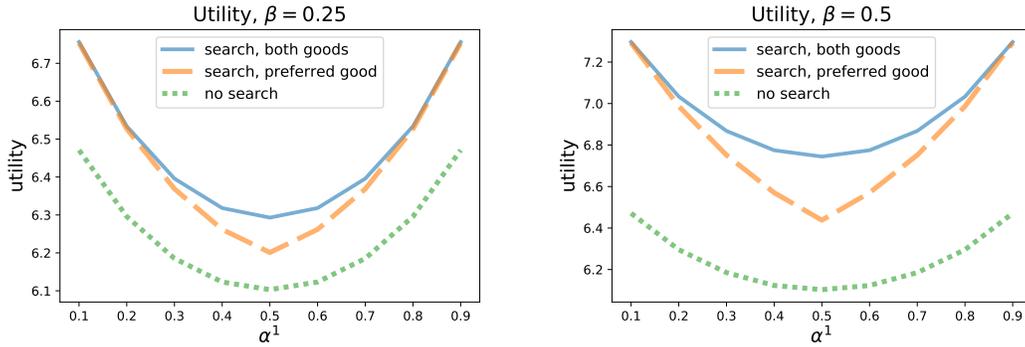
**Table 4. Average Values of BHI within Quarters of its Distribution**

	(1)	(2)	(3)	(4)	(5)
	Consumers Sorted by BHI				
	All	Highest Savings	2 <sup>nd</sup> Highest Savings	2 <sup>nd</sup> Lowest Savings	Lowest Savings
Percent Savings					
Total	0.00	11.75	1.88	-3.74	-10.21
More preferred categories	0.74	13.91	2.69	-3.49	-10.50
Less preferred categories	-2.28	7.17	-0.57	-4.37	-9.60

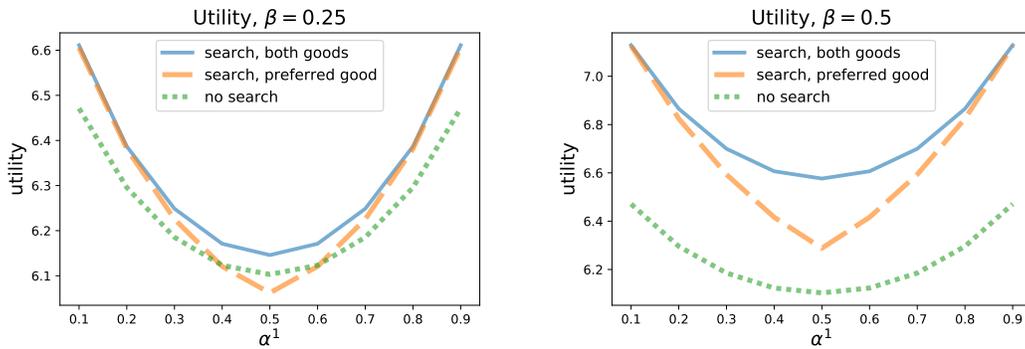
*Notes:* The table displays in the first row the value of the bargain hunting index (BHI), normalized to be 0 on average, and the BHI for consumers with the highest, second-highest, second-lowest, and lowest BHI. For each consumer, we rank his or her purchased categories in terms of the quantity index and divide them into top-half and bottom-half purchased categories (we include only consumers who purchase from at least two categories). We then compute two BHIs for top-half (more preferred) and bottom-half (less preferred) categories. The average values of these two BHIs are presented in the second and third rows of the table, first for all consumers in column (1), and then by quartile-group of the overall BHI in columns (2) through (5).

Figure 1. To Search or Not To Search. Utility under Different Scenarios

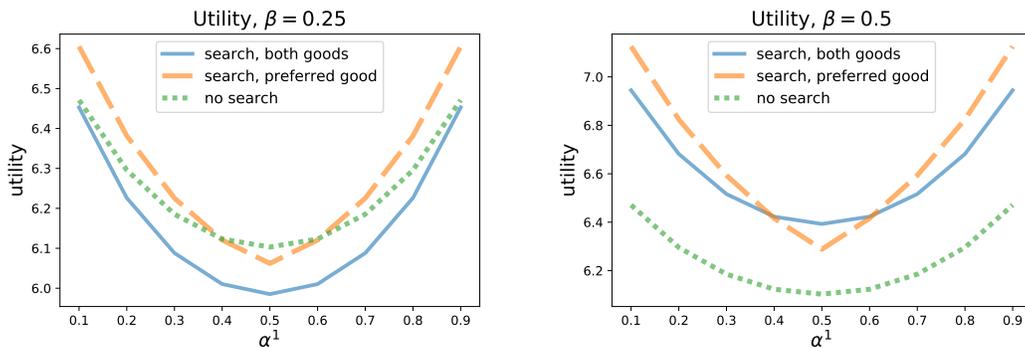
Panel A: No Fixed Search Cost



Panel B: Positive Search Fixed Cost



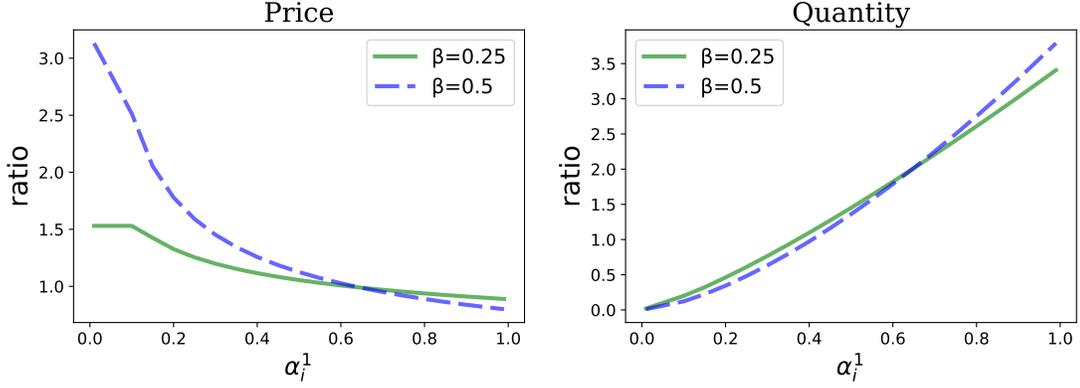
Panel C: Positive Search Fixed Cost, Higher for Two Goods



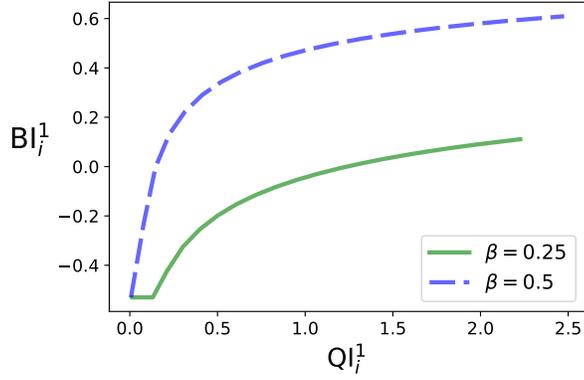
Notes: Each plot depicts utility under three scenarios: No search, search only for the most preferred good (highest  $\alpha^u$ ), and search for both goods. The model parameters are as follows:  $T = 50$ ,  $P_H = 1$ ,  $\eta = 1$ ,  $\mu = 0.5$ ,  $W^0 = 50$ ,  $W^1 = 5$ . Searching efficiency, and relative preference for good 1 ( $\beta$  and  $\alpha^1$ ) vary as indicated in each subplot. In Panel A, the search fixed cost is zero,  $T^0 = 0$ . In panel B,  $T^0 = 5$  when searching for one or two goods. In Panel C,  $T^0 = 5$  when searching for one good, and  $T^0 = 10$  when searching for two goods.

Figure 2. Relative Prices and Quantities

Panel A: Price and Quantity Relative to Average



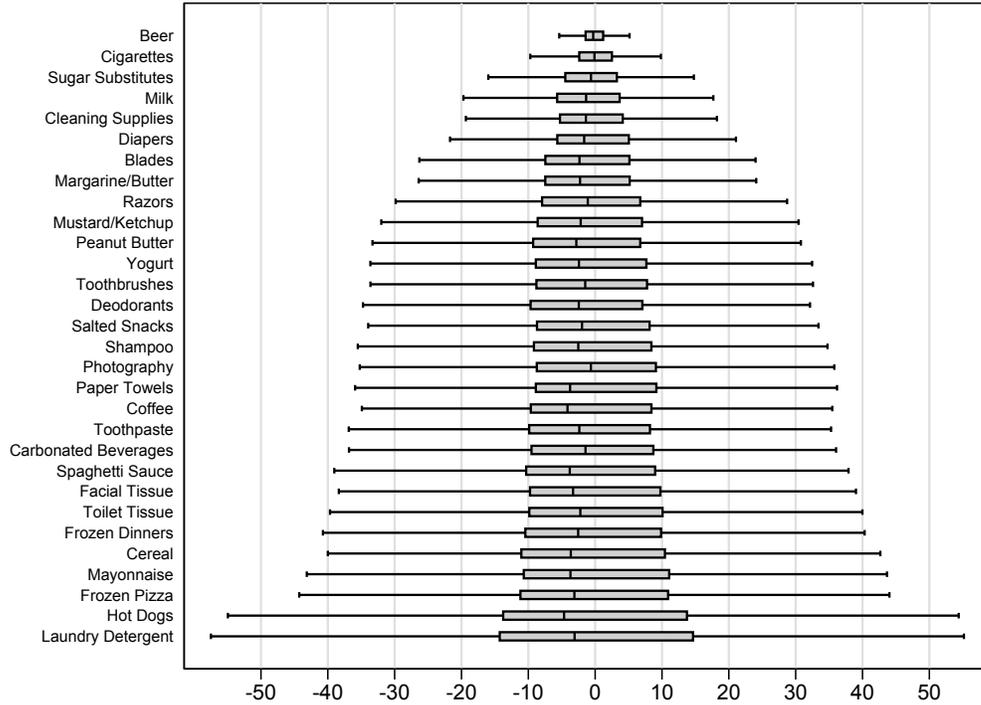
Panel B: Bargain-Hunting and Quantity Indices



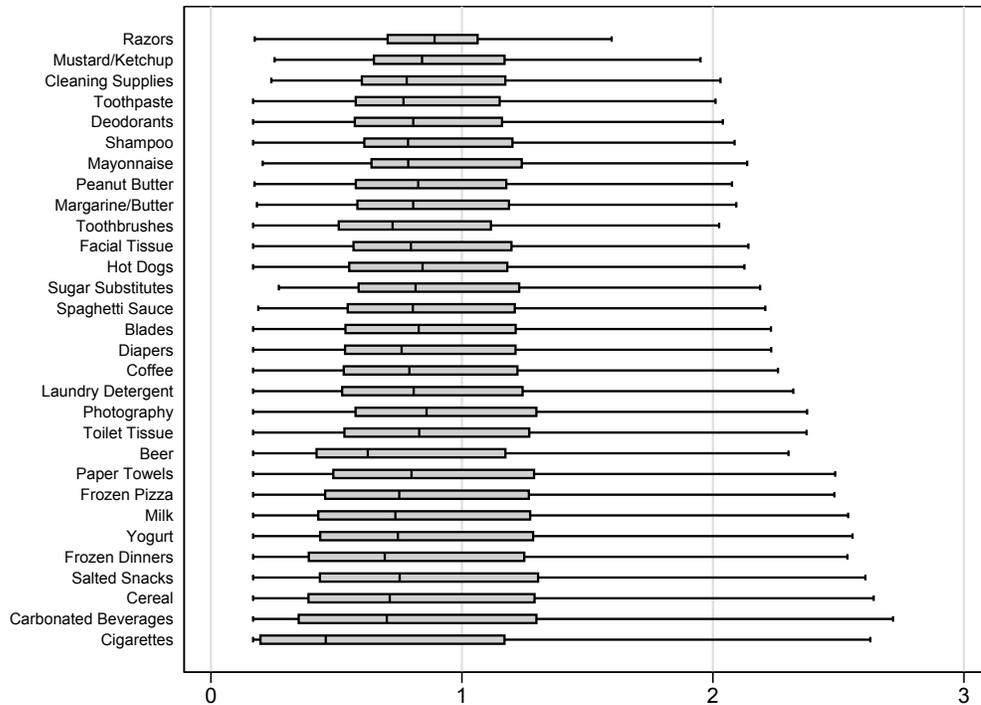
*Notes:* Panel A depicts the relative (to the average) prices paid and quantities consumed of good 1 for individuals with different preferences (indicated on the  $x$ -axis). Panel B depicts the bargain-hunting index,  $BH_i^1$ , for good 1 on the  $y$ -axis and the corresponding quantity index ( $QI_i^1$ ) on the  $x$ -axis for the individuals in Panel A. Each point on the line represents a different individual. A higher  $\alpha_i^1$  results in a higher  $QI_i^1$ , as shown in Panel A.  $BH_i^1 = 1 - (P_i^1 \times Q_i^1) / (\bar{P}^1 \times \bar{Q}^1)$  and  $QI_i^1 = Q_i^1 / \bar{Q}^1$ . We assume a uniform distribution of consumers with  $\alpha_i^1 \in [0, 1]$ . Search efficiency and preference for good 1 ( $\beta$  and  $\alpha^1$ ) vary as indicated in each subplot. The remaining model parameters are as follows:  $T = 50$ ,  $P_H = 1$ ,  $\eta = 1$ ,  $\mu = 0.5$ ,  $W^0 = 50$ ,  $W^1 = 5$ ,  $T^0 = 5$ .

Figure 3. Variation by Category

Panel A: The Bargain-Hunting index



Panel B: The Quantity Index



Notes: Product categories are defined by IRI. Panel A shows variation in the bargain-hunting index (BHI). The left and right borders of each category box depict the 75th and 25th percentiles of the BHI for that category, while the whiskers represent upper and lower adjacent values (outside values not plotted). The categories have been sorted by the interquartile range. Panel B depicts variation in the quantity index by category and is created analogously.

## Online Appendix

### Detailed Solution of the Model

We provide the details of solving the model via a standard Lagrange technique. We suppress the subscript  $i$  that denotes individuals to simplify notation. The Lagrangian is  $L = \alpha^1 \ln Q^1 + \alpha^2 \ln Q^2 + \mu \ln(T - T^Y - T^1 - T^2 - T^0) + \lambda[Y(T^Y) - P^1 Q^1 - P^2 Q^2]$ , and the first order conditions (FOCs) with respect to (wrt.) consumption are:

$$\frac{\alpha^u}{Q^u} = \lambda P^u; u = 1, 2. \quad (9)$$

This implies that  $\frac{Q^1}{Q^2} = \frac{\alpha^1 P^2}{\alpha^2 P^1}$ ; that is, a higher  $\alpha^1$  (higher weight on good 1) increases  $Q^1$  over  $Q^2$ . A higher relative price of good 2 has the same effect. Substituting into the budget constraint, (2), we find that expenditure shares for the two goods are constant:  $P^1 Q^1 = \alpha^1 Y(T^Y)$  and  $P^2 Q^2 = \alpha^2 Y(T^Y)$ .

The FOCs of the Lagrangian wrt.  $T^u$ ,  $u = 1, 2$  are:

$$-\mu(T - T^Y - T^1 - T^2 - T^0)^{-1} - \lambda Q^u \frac{dP^u}{dT^u} = 0; u = 1, 2. \quad (10)$$

Combining the conditions, we find that the marginal gain from search time is equalized across goods:

$$\frac{dP^2}{dT^2} Q^2 = \frac{dP^1}{dT^1} Q^1, \quad \text{or} \quad \frac{\frac{dP^2}{dT^2}}{\frac{dP^1}{dT^1}} = \frac{Q^1}{Q^2}.$$

Given that  $Q^u = \alpha^u / \lambda P^u$ , from FOC (9), and that  $dP^u/dT^u = P^h \beta (T^u + \eta)^{-\beta-1}$ , we

can rewrite the previous expression as:

$$\frac{-\beta(T^2 + \eta)^{-\beta-1}}{-\beta(T^1 + \eta)^{-\beta-1}} = \frac{\alpha^1 (T^2 + \eta)^{-\beta}}{\alpha^2 (T^1 + \eta)^{-\beta}} \quad \text{or} \quad \frac{T^1 + \eta}{T^2 + \eta} = \frac{\alpha^1}{\alpha^2}.$$

That is, relative time allocated to searching for goods is proportional to their relative preferability.

The FOC of the Lagrangian wrt.  $T^Y$  is  $-\mu(T - T^Y - T^1 - T^2 - T^0)^{-1} + \lambda \frac{dY}{dT^Y} = 0$ , and combining this FOC with FOC (10), we obtain  $-Q^u \frac{dP^u}{dT^u} = \frac{dY}{dT^Y}$ . That is, the marginal gain from a unit increase in shopping time is equal to the marginal loss of income.

Substituting for the price derivative, we obtain  $Q^u \frac{dP^u}{dT^u} = Q^u(-\beta P^u (T^u + \eta)^{-1}) = -\beta(P^u Q^u)(T^u + \eta)^{-1}$ , and as  $Q^u P^u = \alpha_i^u Y(T^Y)$ , we find  $\beta \alpha_i^u Y(T^Y)(T^u + \eta)^{-1} = \frac{dY}{dT^Y}$  or  $T^u + \eta = \beta \alpha_i^u \frac{Y(T^Y)}{\frac{dY}{dT^Y}}$ , implying that  $T^u = \beta \alpha_i^u \left( \frac{W^0}{W^1} + T^Y \right) - \eta$  and  $T^1 + T^2 = \beta \left( \frac{W^0}{W^1} + T^Y \right) - 2\eta$ . FOC (9) and the fact that  $Q^u P^u = \alpha_i^u Y(T^Y)$  imply that  $\lambda = 1/Y(T^Y)$ . Given that  $\frac{dY}{dT^Y} = W^1$  and substituting for  $\lambda$ , we can rewrite the FOC wrt.  $T^Y$  as:

$$\mu(T - T^Y - T^1 - T^2 - T^0)^{-1} = \frac{W^1}{W^0 + W^1 T^Y}.$$

Substituting for the value of  $T^1 + T^2$ , we can solve for  $T^Y$  :

$$T^Y = \frac{T - T^0 + 2\eta - (\beta + \mu) \frac{W^0}{W^1}}{1 + \beta + \mu}. \quad (11)$$

Plugging the value of  $T^Y$  into the solution for  $T^u$ , we obtain:

$$T^u = \beta\alpha^u \frac{T - T^0 + 2\eta + \frac{W^0}{W^1}}{1 + \beta + \mu} - \eta. \quad (12)$$

Leisure is

$$T - T^1 - T^2 - T^Y - T^0 = \frac{\mu}{1 + \beta + \mu} \left( T - T^0 + \frac{W^0}{W^1} + 2\eta \right).$$

Assume with no loss of generality that good 1 is the preferred good. Consider a consumer who searches only for prices of good 1, because of the constraint that search time needs to be non-negative is binding for good 2. In this case, optimal work time is

$$T^Y = \frac{T - T^0 + \eta - (\beta\alpha^1 + \mu)\frac{W^0}{W^1}}{1 + \beta\alpha^1 + \mu}. \quad (13)$$

Search time for good 1 is

$$T^1 = \beta\alpha^1 \frac{T - T^0 + \eta + \frac{W^0}{W^1}}{1 + \beta\alpha^1 + \mu} - \eta,$$

and leisure time becomes

$$T - T^1 - T^Y - T^0 = \frac{\mu}{1 + \beta\alpha^1 + \mu} \left( T - T^0 + \frac{W^0}{W^1} + \eta \right).$$

Without search, consumers will pay the higher price for each good, and work and leisure time will be

$$T^Y = \frac{T - \mu \frac{W^0}{W^1}}{1 + \mu}, \quad (14)$$

and

$$T - T^Y = \frac{\mu}{1 + \mu} \left( T + \frac{W^0}{W^1} \right).$$

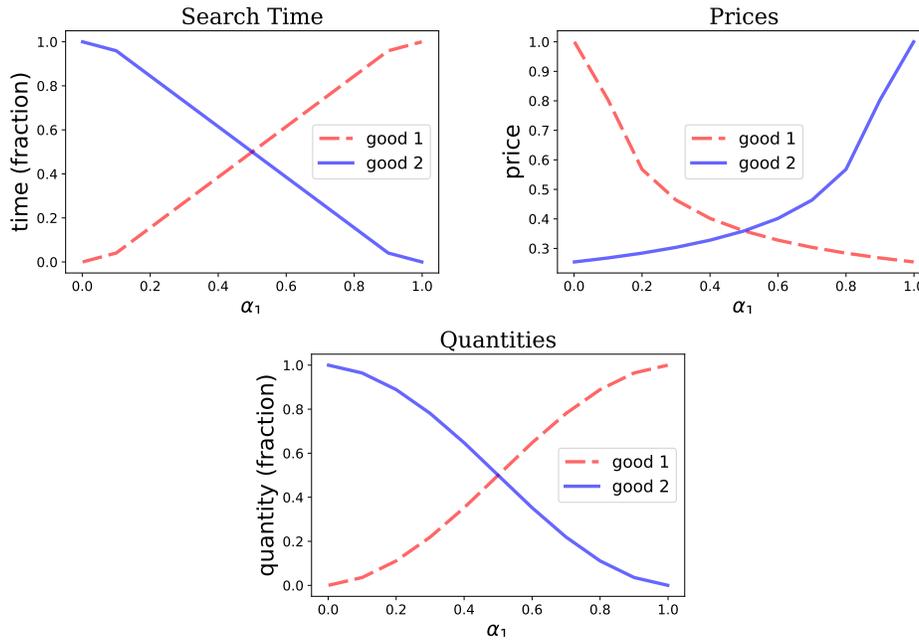
Plugging the full solutions into the utility function for various values of the parameters allows us to determine which of these discrete choices is preferred.

To illustrate the empirical implications of the model, we select certain parameter values and plot optimal search times, prices, and quantities in Figure A.1. We vary the relative preference for good 1, captured by the parameter  $\alpha^1$  ( $i$  subscript omitted), and the efficiency of the search function,  $\beta$ —the higher  $\beta$ , the lower the prices paid for the same level of search. All other parameters are kept constant and are detailed in the notes to the figure. Note the fixed cost of search,  $T^0$ , is set to zero in both cases.

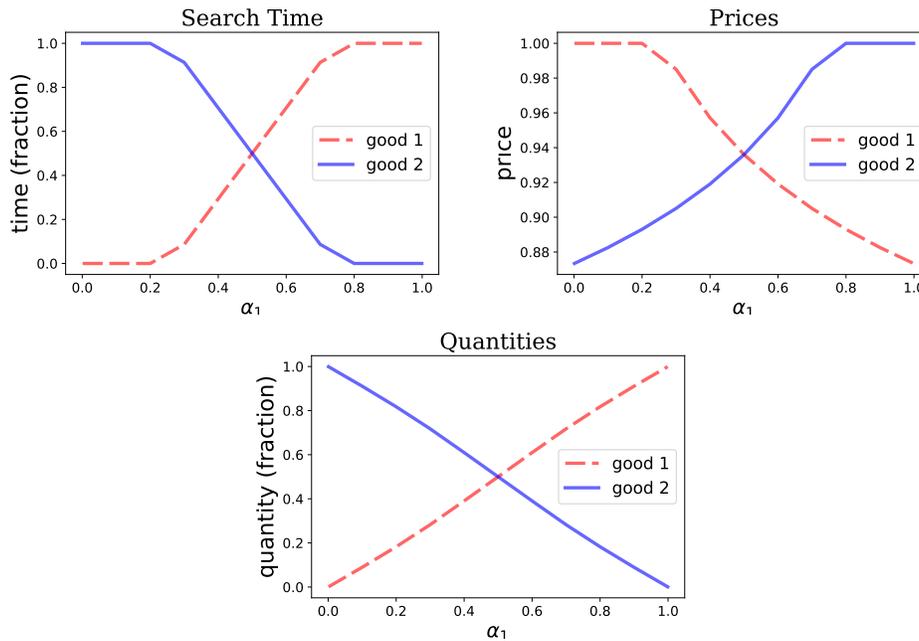
Panel A displays the case of relatively high search efficiency,  $\beta = 0.5$ . Search time for good 1 (2) increases (decreases) with  $\alpha^1$ . The price paid for good 1 declines with  $\alpha^1$  due to the higher search intensity, and the consumer shifts the basket towards higher consumption of good 1 as his or her preference for good 1 increases. *Ceteris paribus*, the model implies an inverse relationship between the prices paid and the quantities consumed of a given good, because consumers vary their search intensity across goods in accordance with their relative preferences. We lower search efficiency in Panel B to illustrate that when search efficiency is relatively low, the consumer optimally chooses not to search for one of the goods even when the search fixed cost is zero ( $\beta = 0.1$  in this case).

Figure A.1. Search Time, Price and Quantity by Preference for Good 1

Panel A: High Search Efficiency



Panel B: Low Search Efficiency



Notes: The figure depicts optimal shopping times, prices and quantities according to the model of Section 2. The model parameters are as follows:  $T = 50$ ,  $P^H = 1$ ,  $\eta = 1$ ,  $\mu = 0.5$ ,  $T^0 = 0$ ,  $W^0 = 50$ ,  $W^1 = 5$ . Search efficiency,  $\beta$ , is 0.5 in Panel A and 0.1 in Panel B. In the plots, search time for good  $u$  is reported as fraction of the total shopping time  $T^u / (T^1 + T^2)$ , and similarly for quantity,  $Q^u / (Q^1 + Q^2)$ .  $\alpha^1$  measures the relative preference for good 1, as  $\alpha^1 + \alpha^2 = 1$ .

## Store Pricing When Consumers Search For Only Some Goods

Sellers rationally differentiate prices across stores and/or over time. Authors, going back to at least Salop and Stiglitz (1977), have constructed models where different prices for the same good across stores persist when some consumers are informed and others are not. Different prices can be rationalized from our consumer model as well. Intuitively, some consumers behave as if they are uninformed about prices because the value of work time (or leisure) is too high to search, while some consumers behave as if they are informed about prices because they rationally search for low prices for all goods they consume. In the literature, price setting when consumers vary in their (overall) search intensity has been shown to imply price differentiation, and it is intuitive that the pattern that we document can also rationalize price differentiation.

In this sub-section, we outline a store-pricing model, which can rationalize price dispersion. Economists have previously developed models for why stores may post different prices for identical goods. For example, Chevalier and Kashyap (2019) assume three types of consumers and two goods (1 and 2) that can be stored. Some consumers have an inelastic demand for good 1, other consumers have an inelastic demand for good 2, and the rest are bargain hunters. This last group will shop for low prices and store goods. Such a model can rationalize why stores have periodical sales.

Kaplan et al. (2019) document price dispersion across stores. Different stores tend to sell different goods at different prices, and store prices are quite persistent. Based on these facts, they develop a model, matching price persistence, where some agents are shoppers

and some are inattentive in order to explain pricing patterns.

We build on these models and suggest a model of price dispersion across stores and goods where some consumers are bargain hunters (searchers or shoppers) only for the goods for which they have relatively strong demand. We will consider this model for the simplest case of two goods.

Consider two goods, indexed by the numbers 1 and 2. Assume stores can set a price  $P_H$ , which is the highest price that a consumer who does not search for that good will pay—for simplicity, we assume this price is constant. For the good a consumer wants in large quantities (his/her preferred good), he or she will search until the marginal value of further search is nil. Assume that consumers who search pay a low price  $P_L^s$ , which differs by store  $s$ . We assume that the price  $P_L^s$  is set competitively such that stores with a higher price provide more amenities. For example, it could be that it takes longer to get to stores with the lower prices due to location (which would literally fit into our framework).

Consumer may search for good 1, good 2, both goods, or not search at all. In this illustration, we assume half of consumers search for good 1 only and half search for good 2 only. When consumers go to the store, they purchase a smaller amount of their less preferred good, if any. A consumer who prefers good 1 searches till he or she finds the lowest price  $P_L^s$  that is consistent with optimal time spent searching, buying an amount  $M_L^s$ . S/he also buys an amount  $M_H$  ( $M_H < M_L^s$ ) of the less preferred good 2 at price  $P_H$ . A consumers who prefers good 2 searches till s/he finds the optimal price for good 2 (symmetric to good 1), and buys a smaller amount of good 1 at the higher price. Further assume that there are a large number of stores so that other stores will not respond to a

given store's change in pricing. A store may set a price  $P_L^s$  for good 1 and  $P_H$  for good 2. The store pays a constant cost  $c^s$  for goods. The store's profit, where the factor reflects that half the potential purchases go to another store, is:

$$\frac{1}{2}[(P_L^s - c^s)M_L^s + (P_H - c^s)M_H] .$$

Due to competition,  $P_L^s$  is set at a minimum that allows a normal profit. An alternative pricing strategy would be to charge  $P_L^s$  for both goods to attract both types of purchases— at any price higher than  $P_L^s$ , consumers will go elsewhere to find their more preferred good. In this case, the store's profit is:

$$\frac{1}{2}[(P_L^s - c^s)M_L^s + (P_L^s - c^s)M_H] + \frac{1}{2}[(P_L^s - c^s)M_H + (P_L^s - c^s)M_L^s] = (P_L^s - c^s)(M_L^s + M_H) .$$

A store cannot charge more than  $P_L^s$  without losing all the purchases of consumers who prefer good 1, and it cannot charge more than  $P_H$  without losing all sales. A store has no incentive to charge less than  $P_H$  unless it lowers the price to  $P_L^s$ .

For a store to differentiate prices the following condition must hold:

$$\frac{1}{2}[(P_L^s - c^s)M_L^s + (P_H - c^s)M_H] > (P_L^s - c^s)(M_L^s + M_H),$$

or

$$(P_H - P_L^s)M_H > (P_L^s - c^s)(M_L^s + M_H).$$

That is, the extra gain from charging a high price for the inelastic demand  $M_H$  outweighs the profit from selling the amount  $M_L^s + M_H$  at a lower price.

## Data

In our dataset, age is reported in categories, and the age distribution by category in January of 2017 is as follows: 22 percent are younger than 45 years old; 25 percent are aged 45 to 54; 22 percent are aged 55 to 64; and 31 percent are 65 or older. Household income is reported by category: 16 percent have income that is less than \$20,000; 22 percent earn \$20,000 to \$35,000; 25 percent earn \$35,000 to \$55,000; 18 percent earn \$55,000 to \$75,000; 11 percent earn \$75,000 to \$100,000; and 8 percent have income that is more than \$100,000. Education categories have the distribution: 5 percent of panelists have not completed high school; 32 percent are high school graduates; 39 percent have some education beyond high school without a college degree, while the rest have graduated from college. Relative to the U.S. population, the IRI sample is somewhat older and poorer, and spending in the IRI categories represents roughly 20 percent of PSID food-at-home expenditure.

**Table A.1. Summary Statistics for Panelist in January of 2007**

	Count	Mean	SD	Min	Max
Years of Education	4,434	13.76	2.01	6	18
Age	4,740	55.34	12.73	21	70
Household Income	4,738	52,537	36,662	5,000	150,000
Old (65+)	4,740	0.31	0.46	0	1
Expenditure (monthly)	4,740	79.47	63.93	5	1,015

Notes: Authors' calculations using all IRI panelist data for January of 2007.

**Table A.2. Summary Statistics for Regressions**

	Count	Mean	SD	Min	Max
Bargain Hunting Index (BHI)	551,438	7.45	9.29	-12	34
BHI (demeaned)	551,438	-0.00	8.72	-27	32
BHI (demeaned), 65+	190,607	0.62	9.21	-27	32
BHI (demeaned), age<65	360,831	-0.33	8.44	-27	32
BHI (demeaned), exp. < median exp.	275,596	0.50	9.78	-27	32
BHI (demeaned), exp. $\geq$ median exp.	275,842	-0.50	7.49	-27	31
Category-Specific BHI	3,728,872	0.00	14.33	-68	63
Category-Specific Quantity Index	3,728,872	0.99	0.74	0	4
Expenditure (monthly)	551,438	69.51	55.32	5	2,281
Old (65+)	551,438	0.35	0.48	0	1
# trips (monthly)	551,438	8.96	6.39	1	126
# stores visited (monthly)	551,438	2.97	2.01	1	34

Notes: Authors' calculations using all IRI panelist data from 2003 through 2012. The BHI computation is described in equation (5). The index measures how much a consumer saves (positive values), in percent, or overpays (negative values) relative to buying his or her consumption bundle at average prices. The BHI is broken up by age group and expenditure group. The category-specific BHI is described in equation (5) and focuses on savings in a specific category. The category-specific quantity index, which measures whether a consumer purchases more or less of that category than does the average consumer, is computed according to equation (7). The other variables are used in our regressions: (1) Expenditure is total dollars spent in a given month by a panelist in IRI transactions; (2) Old (65+) is a dummy variable for whether consumers are 65 or older; (3) # trips to store (monthly) is the total number of trips to stores by a panelist in a given month; (4) # stores visited (monthly) is the number of different stores that a consumer visits in a given month.

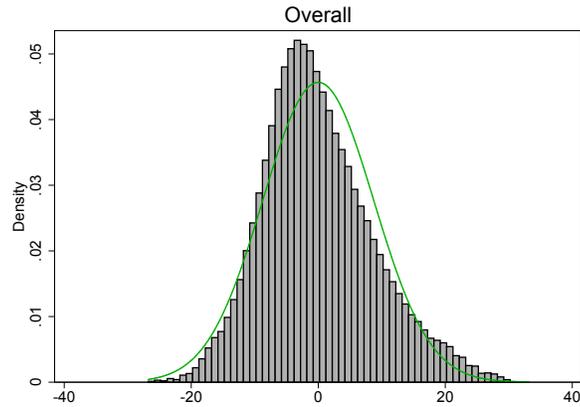
## **Additional Figures**

In Figure A.2, we use a histogram to display the dispersion of the (demeaned) overall BHI. The BHI is slightly leptokurtic (kurtosis is 3.3) and skewed to the right (skewness is .43). The bottom two panels split the sample into 65-plus and younger panelists, and into panelists with below- and above-median expenditure in a given period. As our model predicts, the older individuals pay lower prices on average than do the younger panelists, and the poorer panelists (as proxied by expenditure) also pay relatively lower prices.

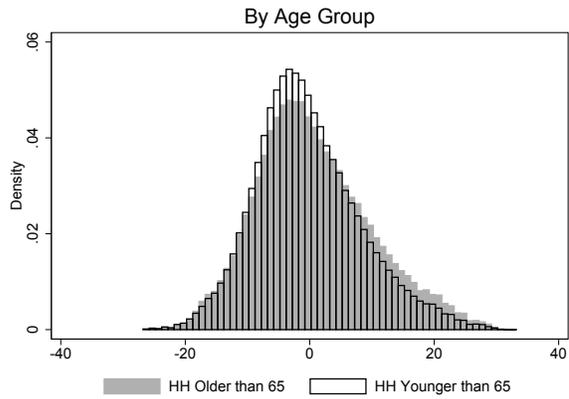
Figure A.3 depicts histograms for the overall BHI and the store BHI. The histograms indicate savings from both store selection (consumers' purchasing products in stores where they are relatively cheaper) and the timing of purchases within a given store.

Figure A.2. The bargain-hunting index

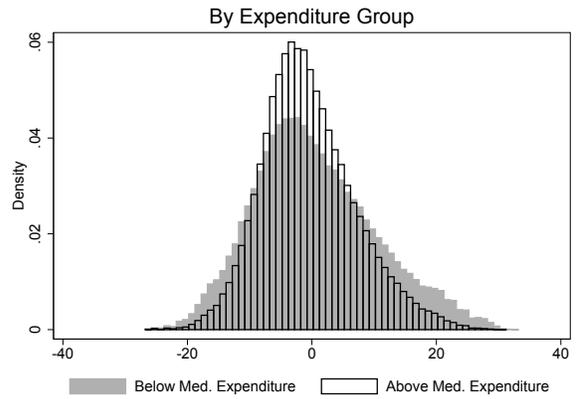
Panel A: Overall



Panel B: By Age Group



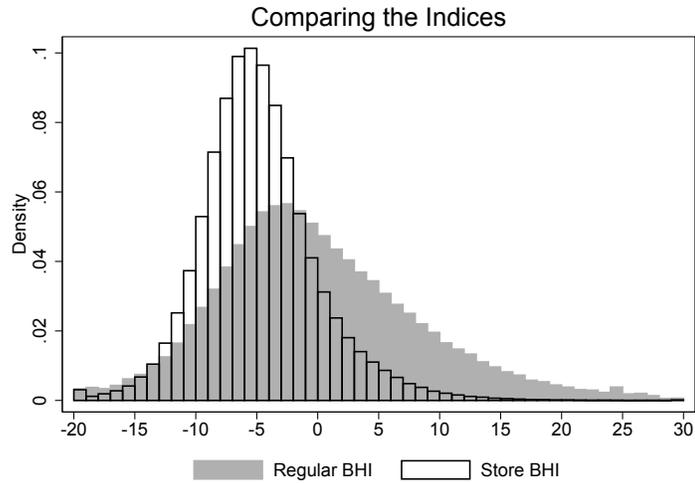
Panel C: By Expenditure Group



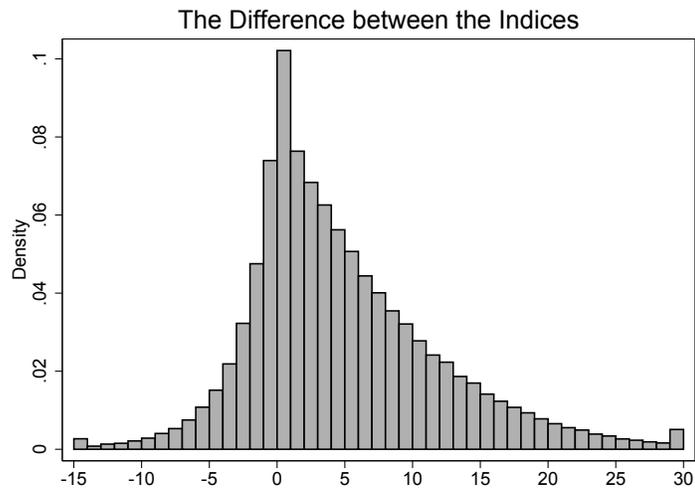
*Notes:* The BHI index shows how much a consumer saves, in percentages, relative to the counterfactual of buying his or her consumption bundle at average prices. The BHI index has been normalized to have a mean of 0 every month-year by market. *Source:* IRI, all panelist data from 2003 through 2012.

**Figure A.3. The BHI vs. the Store BHI**

**Panel A: Comparing the Indices**

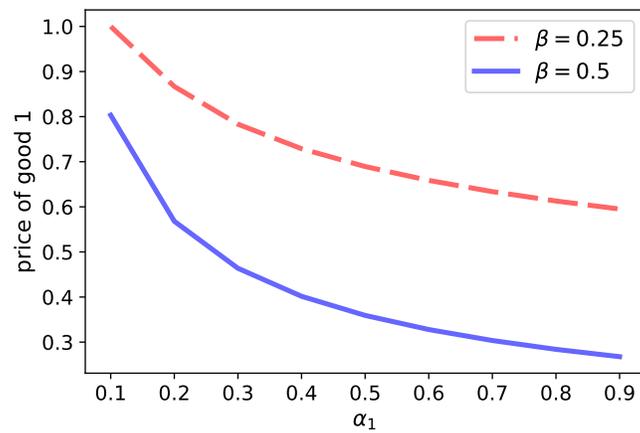


**Panel B: The Difference between the Indices**



*Notes:* The regular BHI index is defined in equation (5) and represents how much a consumer saves relative to buying at average prices across stores. The store BHI is defined in equation (8) and measures how much a consumer would save if he or she paid average prices in the store relative to buying the consumption bundle at average prices across all stores. Panel B plots the distribution of the difference between the indices (individual by individual). *Source:* IRI, all panelist data from 2003 through 2007.

Figure A.4. Search Efficiency, Average Prices, and Price Dispersion



*Notes:* This figure illustrates that price dispersion may not decrease with improvements in search efficiency. Mean (SD) [CV] of prices with  $\beta = 0.25 : 0.73$  (0.13) [0.18]. Mean (SD) [CV] of prices with  $\beta = 0.50 : 0.42$  (0.16) [0.38].

## Further Robustness

To test whether our results are robust across categories, we estimate the regression separately for each category  $c$ , using the average category-specific quantity index as an instrument. The data in each regression form an individual  $\times$  time panel, and all coefficients, including dummies, can take different values for the different categories.

$$\text{BHI}_{i,t}^c = \mu^c + \gamma_{m,t}^c + \beta^c \text{QI}_{i,t}^c + X_{i,t} \alpha^c + \epsilon_{i,t}$$

Table A.3 in the online appendix shows the regressions category-by-category.<sup>26</sup> We will not discuss each category in detail, but together the results reveal that our main qualitative result is remarkably robust—the coefficient to the quantity index is positive and highly significant in almost all categories. The exceptions are beer, for which the estimated coefficient is virtually 0, and cigarettes for which the coefficient is negative and insignificantly different from 0. These two categories are the ones with the lowest price dispersion and, therefore, the lowest return to bargain hunting. The size of the coefficients to the quantity indices vary by category, with the smallest coefficients found for categories with relatively less price variation at the UPC level. In particular, all categories with a coefficient less than unity are among those with the lowest price variation. The largest coefficient is for hot dogs, the category with the second highest price variation. In sum, while the variation in the size of the coefficients is not one-to-one with price dispersion, the variation in the coefficients reflects the potential gains from bargain hunting as captured

---

<sup>26</sup>The quantity indices in these regressions have been standardized to have a mean 0 and a standard deviation of 1 by category for an easier comparison across the 30 regressions.

by the price variation.

Table A.4 shows that results are qualitatively similar to those obtained at the monthly frequency when aggregating purchases and averaging prices to the quarterly frequency, which supports the causal interpretation of the results.

**Table A.3. The BHI and the QI by Category. Separate IV-Regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Quantity Index	0.01 (0.04)	0.78** (0.35)	1.92*** (0.13)	-0.23* (0.13)	1.87*** (0.26)	3.47*** (0.23)	5.33*** (0.65)	1.45** (0.59)	3.00*** (0.36)	1.69*** (0.30)
Log. Expenditure	-0.14*** (0.03)	-0.96*** (0.23)	-1.85*** (0.09)	0.16 (0.13)	-1.21*** (0.14)	-2.10*** (0.13)	-2.10*** (0.19)	-1.45*** (0.33)	-1.78*** (0.15)	-1.75*** (0.17)
Old (65+)	-0.11** (0.05)	0.35 (0.35)	0.64*** (0.14)	-0.39 (0.27)	1.34*** (0.21)	-0.68*** (0.18)	0.78** (0.31)	1.30*** (0.46)	0.31 (0.23)	0.70*** (0.21)
Observations	86504	17228	354994	25430	153068	301697	42551	11575	110072	164925
F-test excl. inst.	65883	3705	199769	24933	45308	115289	9021	1705	37443	48932
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Quantity Index	1.65*** (0.26)	1.24*** (0.27)	6.00*** (0.39)	3.23*** (0.35)	2.07*** (0.24)	6.82*** (0.33)	0.59*** (0.09)	5.33*** (0.48)	4.03*** (0.24)	4.45*** (0.34)
Log. Expenditure	-1.72*** (0.15)	-0.52*** (0.11)	-2.77*** (0.17)	-3.34*** (0.18)	-0.56*** (0.11)	-1.81*** (0.12)	-0.38*** (0.06)	-1.37*** (0.15)	-2.01*** (0.13)	-1.20*** (0.14)
Old (65+)	0.64*** (0.21)	0.03 (0.17)	0.28 (0.25)	0.57** (0.28)	0.17 (0.18)	0.65*** (0.20)	0.04 (0.11)	0.24 (0.21)	-0.37* (0.20)	0.85*** (0.24)
Observations	159206	52823	112348	146606	138726	120653	423015	73911	115133	101725
F-test excl. inst.	30098	9631	18727	39901	50813	23369	297080	8467	52571	24106
	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Quantity Index	4.19*** (0.83)	1.65*** (0.52)	2.21*** (0.17)	3.07*** (0.41)	3.87*** (0.30)	1.27*** (0.26)	4.85*** (0.25)	3.41*** (0.88)	5.25*** (0.39)	1.03*** (0.16)
Log. Expenditure	-1.87*** (0.42)	-1.66*** (0.53)	-1.08*** (0.10)	-1.31*** (0.17)	-1.38*** (0.13)	-0.83*** (0.18)	-2.61*** (0.12)	-0.66** (0.27)	-2.01*** (0.15)	-0.67*** (0.09)
Old (65+)	0.58 (0.68)	0.62 (0.92)	0.78*** (0.15)	0.09 (0.29)	1.77*** (0.22)	0.67*** (0.25)	-0.10 (0.17)	-0.08 (0.43)	0.55** (0.26)	0.70*** (0.17)
Observations	4752	2159	345689	41029	134273	20863	199147	21116	67796	219417
F-test excl. inst.	870	425	111775	6322	26163	7575	54252	3235	13165	79723
Month-Year × Market FE	Yes									

*Notes:* Regression:  $BHI_{i,t}^c = \gamma_{m,t}^c + \beta^c QI_{i,t}^c + X_{i,t} \alpha^c + \epsilon_{i,t}$ , estimated category by category. The quantity index,  $QI_{i,t}^c$ , is standardized by category (mean 0, sd 1) for easier interpretation. All regressions include market × month FE and are estimated by IV. Our instrument is the average category-specific quantity index, defined more precisely in the notes to Table 2. Standard errors clustered by panelist. \*\*\* (\*\*) [\*] significant at the 1 (5) [10] percent level.

Categories as follows: (1) beer, (2) blades, (3) carbonated beverages, (4) cigarettes, (5) coffee, (6) cold cereal, (7) deodorants, (8) diapers, (9) facial tissue, (10) frozen dinners, (11) frozen pizza, (12) cleaning supplies, (13) hot dogs, (14) laundry detergent, (15) margarine/butter, (16) mayonnaise, (17) milk, (18) mustard/ketchup, (19) paper towels, (20) peanut butter, (21) photography, (22) razors, (23) salted snacks, (24) shampoo, (25) spaghetti sauce, (26) sugar substitutes, (27) toilet tissue, (28) toothbrushes, (29) toothpaste, (30) yogurt.

**Table A.4. Rational Inattention. Pooled Regressions. Quarterly Frequency**

	(1)	(2)	(3)	(4)
	All Years	All Years	Pre-2008	Pre-2008
	BHI	BHI	BHI	Store BHI
Quantity Index	1.70*** (0.06)	1.64*** (0.06)	1.65*** (0.06)	0.26*** (0.02)
Log. Expenditure	-1.04*** (0.06)	-1.31*** (0.06)	-1.43*** (0.06)	-0.43*** (0.03)
Old (65+)	0.51*** (0.12)	0.15 (0.11)	-0.11 (0.12)	-0.08* (0.05)
# stores visited (quarterly)		0.33*** (0.02)	0.53*** (0.02)	0.17*** (0.01)
# trips (quarterly)		0.03*** (0.00)	0.03*** (0.00)	0.01*** (0.00)
Quarter-Year $\times$ Market FE	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	Yes	Yes
Observations	1600439	1600439	889316	889316
Adj. R-squared	0.02	0.03	0.04	0.01
F-test excl. inst.	873232	870518	519741	519741

*Notes:* Expenditure is aggregated at the quarterly level, and store-average prices are calculated at the same frequency. Regression:  $BHI_{i,t}^c = \nu_c + \gamma_{m,t} + \beta QI_{i,t}^c + X_{i,t} \alpha + \epsilon_{it}$ , where  $BHI_{i,t}^c$  is the category-specific bargain-hunting index for individual  $i$  in quarter  $t$ ,  $\nu_c$  denotes category fixed effects.  $\gamma_{m,t}$  is a market  $\times$  month FE,  $X$  is a vector of regressors, and  $QI_{i,t}^c$  is the quantity index described in equation (7). The quantity index, which measures whether a consumer purchases more or less of a category than the average consumer, is standardized (mean 0, sd 1) for easier interpretation. In column (4),  $BHI_{i,t}^c$  is replaced by a category-specific store BHI,  $BHI_{i,t}^{c,s}$ . All regressions are estimated by IV, using the average category-specific quantity index as an instrument, adapted to the quarterly frequency and defined more precisely in the notes to Table 2. Standard errors clustered by individual. \*\*\* (\*\*) [\*] significant at the 1 (5) [10] percent level.