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

Chloe Lee
Federal Reserve Bank of Boston

Wen Long
Barclays Capital

María José Luengo-Prado
Federal Reserve Bank of Boston

Bent E. Sørensen
University of Houston and CEPR

January 27, 2019

Abstract

A model of the allocation of time between work, leisure, and price-search for different goods predicts that consumers spend relatively more time searching for better prices of goods of which they consume relatively more. Using scanner data, we confirm empirically that consumers pay lower (higher) prices for goods that they buy more (less) of than other consumers.

JEL classification: E21, D12

Keywords: Consumption; Consumer Search; Time Use

*Lee: Federal Reserve Bank of Boston, 600 Atlantic Av., Boston, MA, 02210, email: chloe.lee@bos.frb.org; Long: Barclays Capital, 745 7th Ave, New York, NY, 10019, email: longwen1983@gmail.com; Luengo-Prado: Federal Reserve Bank of Boston, 600 Atlantic Av., 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 spent finding lower prices with the value of time 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 time in order to buy Coke at relatively low prices than to buy Budweiser at relatively low prices. Using detailed shopping information from the IRI Academic Dataset, we confirm this prediction. Our results are relevant for understanding consumer rationality. They are also relevant for optimal store pricing and for understanding how consumers can efficiently self-insure by obtaining lower prices for preferred goods in the face of employment shocks.

The 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 constitute 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 different stores when they shop.[1] Nevo and Wong (2018), 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.[2] Nevo and Wong (2018) 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 (2011) 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 during 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.

[1] 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.

[2] Griffith et al. (2009) summarize 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.

Kaplan et al. (2016) 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 less important (for them) goods.

Following Aguiar and Hurst (2007), we construct for each consumer a bargain hunting index (BHI) which measures the price 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 tiny fraction of UPCs.[3]

Using the category BHI, we ask whether consumers who purchase more units than other consumers in a particular category pay relatively lower prices in that category. We find that they do, and the finding is very robust: If we regress prices on quantities category by category, the pattern holds for all categories. This result holds in regressions with individual-specific fixed effects, which implies that consumers who increase their consumption in a category will find lower prices for the goods in that category. 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.

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 that consumers pay relatively less for goods of which they purchase relatively more as they optimally search relatively more (or drive farther) for better prices of goods which 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. In the appendix, we outline a model of store price setting which is consistent with a consumer model where consumers search only for low prices for some goods.

[3]Figure 3 lists the included categories.

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 obtain lower prices for goods for which he or she spends time searching for better prices and the consumer will rationally choose not to search, to search for one good, or to search for two goods, based on the utility of each choice. We formulate a simple model with differentiable search and utility functions but with a discrete choice of 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.

2.1 Searching for best prices of two goods

Consider a consumer who derives utility as summarized by the objective function

$$\max_{C_1,C_2,T_Y,T_1,T_2} \quad \alpha_1 \ln(C_1) + \alpha_2 \ln(C_2) + \mu \ln(T - T_Y - T_1 - T_2 - 2\,T_0), \tag{1}$$

$$\text{subject to:} \qquad P_1 C_1 + P_2 C_2 = Y(T_Y), \tag{2}$$

where T_Y is the time devoted to income-generating activities (work). T_i ($i = 1, 2$) is (continuous) time spent searching for good i and T_0 is a search fixed cost for each good. "Search time" refers to time used in order to obtain lower prices and not just price discovery. For example, T_0 could be the time spent reading advertisements, while T_i could be the time spent driving to the store where good i is purchased. T is the total time endowment, and $T - T_Y - T_1 - T_2 - 2\,T_0$ is leisure. Solutions with negative time spent in any activity are not valid but for ease of exposition, we do not explicitly write down Kuhn-Tucker conditions.

Time spent searching results in lower prices according to the function $P_i = P_h(T_i + \eta)^{-\beta}$, with $\beta > 0$. The larger β is, the more searching lowers the price paid. P_h is $\frac{P_H}{\eta^{-\beta}}$, so that zero search time corresponds to the highest price, P_H, charged by stores. The marginal effect of an additional unit of search is $\frac{dP_i}{dT_i} = -\beta(T_i + \eta)^{-\beta-1} = -\beta P_i(T_i + \eta)^{-1}$.

Income is a linear function of time spent working, with wage rate W_1, and non-wage income, W_0: $Y(T_Y) = W_0 + W_1 T_Y$. C_1 and C_2 are the purchased quantities of good 1 and good 2, and α_1/α_2 is the preference for good 1 over good 2, with $\alpha_1 + \alpha_2 = 1$.

The Lagrangian is $L = \alpha_1 \ln(C_1) + \alpha_2 \ln(C_2) + \mu \ln(T - T_Y - T_1 - T_2 - 2T_0) + \lambda[Y(T_Y) - P_1 C_1 - P_2 C_2]$, and the first order conditions (FOCs) with respect to (wrt) consumption are:

$$\frac{\alpha_i}{C_i} = \lambda P_i\,;\; i = 1, 2\,. \tag{3}$$

This implies that $\frac{C_1}{C_2} = \frac{\alpha_1 P_2}{\alpha_2 P_1}$; that is, a higher α_1 (higher weight on good 1) increases C_1 over C_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: $P_1C_1 = \alpha_1 Y(T_Y)$ and $P_2C_2 = \alpha_2 Y(T_Y)$.

The FOCs of the Lagrangian wrt T_i, $i = 1, 2$ are:

$$-\mu(T - T_Y - T_1 - T_2 - 2\,T_0)^{-1} - \lambda C_i \frac{dP_i}{dT_i} = 0;\; i = 1, 2\,. \tag{4}$$

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

$$\frac{dP_2}{dT_2} C_2 = \frac{dP_1}{dT_1} C_1, \quad \text{or} \quad \frac{\frac{dP_2}{dT_2}}{\frac{dP_1}{dT_1}} = \frac{C_1}{C_2}.$$

Given that $C_i = \alpha_i/\lambda P_i$, from FOC (3), and that $dP_i/dT_i = P_h \beta(T_i + \eta)^{-\beta-1}$, we can rewrite the previous expression as:

$$\frac{-\beta(T_2 + \eta)^{-\beta-1}}{-\beta(T_1 + \eta)^{-\beta-1}} = \frac{\alpha_1}{\alpha_2} \frac{(T_2 + \eta)^{-\beta}}{(T_1 + \eta)^{-\beta}} \quad \text{or} \quad \frac{T_1 + \eta}{T_2 + \eta} = \frac{\alpha_1}{\alpha_2}\,.$$

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

The FOC of the Lagrangian wrt T_Y is $-\mu(T - T_Y - T_1 - T_2 - 2\,T_0)^{-1} + \lambda \frac{dY}{dT_Y} = 0$, and combining this FOC with FOC (4), we obtain $-C_i \frac{dP_i}{dT_i} = \frac{dY}{dT_Y}$. 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 $C_i \frac{dP_i}{dT_i} = C_i(-\beta P_i (T_i + \eta)^{-1}) = -\beta(P_i C_i)(T_i + \eta)^{-1}$, and since $C_i P_i = \alpha_i Y(T_Y)$, we find $\beta \alpha_i Y(T_Y)(T_i + \eta)^{-1} = \frac{dY}{dT_Y}$ or $T_i + \eta = \beta \alpha_i \frac{Y(T_Y)}{\frac{dY}{dT_Y}}$, implying that $T_i = \beta \alpha_i \left(\frac{W_0}{W_1} + T_Y\right) - \eta$ and $T_1 + T_2 = \beta \left(\frac{W_0}{W_1} + T_Y\right) - 2\eta$. FOC (3) and the fact that $C_i P_i = \alpha_i Y(T_Y)$ imply that $\lambda = 1/Y(T_Y)$. Given that $\frac{dY}{dT_Y} = W_1$ and substituting for λ, we can rewrite the FOC wrt T_Y as:

$$\mu(T - T_Y - T_1 - T_2 - 2\,T_0)^{-1} = \frac{W_1}{W_0 + W_1 T_Y}.$$

Substituting for the value of $T_1 + T_2$, we can solve for T_Y :

$$T_Y = \frac{T - 2\,T_0 + 2\eta - (\beta + \mu)\frac{W_0}{W_1}}{1 + \beta + \mu}\ . \tag{5}$$

Work time increases in total time available and in wages (W_1), and it decreases in nonwage income (W_0), price-search efficiency (β), and leisure preference (μ).[4] Plugging the value of T_Y into the solution for T_i, we obtain:

$$T_i = \beta\alpha_i \frac{T - 2\,T_0 + 2\eta + \frac{W_0}{W_1}}{1 + \beta + \mu} - \eta. \tag{6}$$

Relative time spent searching 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. Search time decreases with wages and leisure preference. Leisure is:

$$T - T_1 - T_2 - T_Y - 2\,T_0 = \frac{\mu}{1 + \beta + \mu}\left(T - 2\,T_0 + \frac{W_0}{W_1} + 2\eta\right).$$

For a fixed W_0, a larger wage, W_1, implies more work time, greater income, and less

[4] We assume that the parameters and W_0 and W_1 are such that T_Y is non-negative.

search time, so the "wealthy" (in terms of labor income) pay more. Work time decreases with the ratio W_0/W_1, and T_i increases, so retirees are predicted to pay lower prices. The higher the value of η, the less time is spent searching for low prices and the more time is spend on work and leisure, this reflects that the marginal return to searching is declining with η. For simplicity, assume W_0/W_1 is such that work time is 0 for retirees. The unconstrained solution may involve negative search time for, e.g., the least preferred good, in which case the solution will be the one where there is no search for that good.

2.2 Search for one good

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. The solution will have the same first order conditions (with $T_2 = 0$) up till equation (5). This equation will now take the form

$$T_Y = \frac{T - T_0 + \eta - (\beta\alpha_1 + \mu)\frac{W_0}{W_1}}{1 + \beta\alpha_1 + \mu} \ . \tag{7}$$

Search time for good 1 will be

$$T_1 = \beta\alpha_1 \frac{T - T_0 + \eta + \frac{W_0}{W_1}}{1 + \beta\alpha_1 + \mu} - \eta$$

and leisure

$$T - T_1 - T_Y - T_0 = \frac{\mu}{1 + \beta\alpha_1 + \mu}\left(T - T_0 + \frac{W_0}{W_1} + \eta\right).$$

2.3 No Search

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

$$T_Y = \frac{T - \mu\frac{W_0}{W_1}}{1 + \mu}\ , \tag{8}$$

and

$$T - T_Y = \frac{\mu}{1 + \mu}\left(T + \frac{W_0}{W_1}\right),$$

which can be plugged into the utility function together with the optimal purchased amounts in order to evaluate the utility of this choice.

2.4 Empirical Implications

To illustrate the empirical implications of the model, we select certain parameter values and plot optimal search times, prices, and quantities for each good in Figure 1. We vary the preference for good 1, captured by the parameter α_1, and the efficiency of the search function, β (the higher β, 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; in particular, the fixed cost of search, T_0, is set to zero in both cases.

We start with a relatively high search efficiency, $\beta = 0.5$, in Panel A. Search time for good 1 (2) increases (decreases) with α_1. The price paid for good 1 (2) declines with α_1 due to the higher search intensity, and the consumer shifts the basket towards higher (lower) consumption of good 1 (2) as his or her preference for good 1 increases (decreases). *Ceteris paribus*, the model implies an inverse relationship between the prices paid and the quantities consumed of a given good—consumers vary their search intensity across goods in accordance with their relative preferences. In Panel B, we lower search efficiency, $\beta = 0.1$, to illustrate that when search efficiency is relatively low the consumer optimally chooses not to search for one of the goods, even if the search fixed cost is zero.

We illustrate the importance of search fixed costs in Figure 2. The figure depicts (restricted) utility under three scenarios: (1) the consumer does not search for better prices at all, (2) the consumer searches only for his or her most preferred good (that with the highest α_i), 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 vary the level of the search fixed cost ($T_0 = 0$, $T_0 = 5$ "low," and $T_0 = 10$ "high"), the level of search efficiency ($\beta = 0.25$ or $\beta = 0.5$), and the relative preferences for good 1, α_1.

When the consumer cares for both goods equally ($\alpha_1 = 0.5$) and the search fixed cost is low and/or search efficiency is high, the consumer is better off spending time searching for lower prices of both goods. Conversely, if fixed costs are high and/or search efficiency is low, the consumer optimally decides not to search at all. 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 summary, the model predicts an inverse relationship between prices and quantities when differential preferences for goods leads 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.

2.5 Store Pricing

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 also can rationalize price differentiation. We, therefore, limit ourselves to outlining, in the appendix, a simple model of price differentiation in which each agent searches for one good but not for the other (which is a possible optimal outcome in our consumer model).

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. 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. 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 since they can be linked to the consumer panel. For regressions on overall expenditure, we drop 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.[5]

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 for 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 and

[5]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.)

stores in the IRI dataset constitutes 22 to 28 percent of food-at-home expenditure.[6]

3.2 Data Construction

Similarly to Aguiar and Hurst (2007), we define the average price of a UPC item, u, in a given market, m, and month, t, as a quantity-weighted average of individual transaction, k, prices for that specific product (u) in that market (m) and month (t) . A transaction in our analysis is the purchase of a given UPC/good during a visit to a store.[7] The average price is

$$\overline{p}^{u,m,t} = \frac{\sum_{k\in u,m,t} q_k^{u,m,t} p_k^{u,m,t}}{\sum_{k\in u,m,t} q_k^{u,m,t}},$$

where q_k is the quantity purchased in transaction k (involving UPC 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 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 result in less precise estimates.[8]

[6] The lower number does not adjust for income differences in the two samples, whereas the higher number does.

[7] One visit to a store usually comprises many transactions.

[8] One may relate, say, average weekly prices (relative to the weekly store-average) to quarterly purchased amounts, as we did in an earlier version of this paper. The qualitative conclusions are similar, and

We define a 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:

$$\text{BHI}_{i,t} = \left(1 - \frac{\text{Actual Exp}_{i,t}}{\text{Hypo Exp}_{i,t}}\right) \times 100 = \left(1 - \frac{\sum_{k=1}^{N_i^t} p_{i,k}^{u,m,t} \times q_{i,k}^{u,m,t}}{\sum_{k=1}^{N_i^t} \overline{p}^{u,m,t} \times q_{i,k}^{u,m,t}}\right) \times 100\,, \qquad (9)$$

where i is a consumer who purchases products in market m, and we aggregate expenditure to the monthly frequency t. N_i^t is the total number of transactions of consumer i. 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 purchase of a good by a household (transaction) in a month, we use the exact price of the good (identified by its UPC, $p_{i,k}^{u,m,t}$) to calculate actual expenditure. Given the consumer's consumption bundle, hypothetical expenditure is measured using the store-average price ($\overline{p}^{u,m,t}$) of the good in the same market and month. Expenditure is aggregated over a month in order to avoid a large number of zero observations at higher frequencies. A higher BHI means saving more (paying less) relative to the store-average prices given the household's consumption bundle. We also calculate the number of shopping trips for each individual in a given month, and the average number of (different) stores visited per month.

we prefer to limit confusion by aggregating quantities over a month to match the averaging frequency for prices.

Table 1 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 likely reflects that panelists in our sample are, on average, older than the typical population.[9] Going forward, we demean the BHI to 0 each period.[10]

Our main focus is on selective bargain hunting (that is, whether consumers devote relatively more time to searching for lower prices for goods they prefer). To test for such a pattern, we construct (1) a bargain hunting index by individual and category in each month, and (2) a quantity index by category that measures whether a consumer buys relatively more or less of that category to proxy for his or her preferences. We could construct groups of UPCs ourselves, but because there is no obvious way of doing this and an arbitrary choice of categories would open up a scope for data mining, we utilize the categories as defined by IRI.

Let c denote a category. A BHI by category for a given consumer i in period t, $\text{BHI}^c_{i,t}$, is computed similarly to the overall BHI, except that only transactions involving products

[9]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.

[10]This is not strictly necessary for the regression analysis, because we include period (year×month) fixed effects.

in a given category are added up:

$$\text{BHI}^c_{i,t} = \left(1 - \frac{\sum_{k\in c} p^{u,m,t}_{i,k} \times q^{u,m,t}_{i,k}}{\sum_{k\in c} \overline{p}^{u,m,t} \times q^{u,m,t}_{i,k}}\right) \times 100\,. \tag{10}$$

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. 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—in our pooled analysis, we include category fixed effects, when relevant, to take this into account. 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, $\text{QI}^c_{i,t}$, 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:

$$\text{QI}^c_{i,t} = \frac{\sum_{k\in c} \overline{p}^{u,m,t} \times q^{u,m,t}_{i,k}}{(\sum_{j\in J^m_t} \sum_{k\in c} \overline{p}^{u,m,t} \times q^{u,m,t}_{j,k})/J^m_t}\,, \tag{11}$$

where J^m_t is the number of consumers in the panel in market m in period t. 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.

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 is (roughly) 1 for each category.

To further explore how consumers save money, we compute for each consumer an alternative "store-price BHI," $\text{BHI}^s_{i,t}$, that computes the value of consumer i's basket using the average price of each UPC in a given month *in the store*, s, where the item was purchased, $\overline{p}_s^{u,m,t}$. To compute this average price, we use the store-level dataset.[11]

$$\text{BHI}^s_{i,t} = \left(1 - \frac{\sum_{k=1}^{N_i^w} \overline{p}_s^{u,m,t} \times q_{i,k}^{u,m,t}}{\sum_{k=1}^{N_i^w} \overline{p}^{u,m,t} \times q_{i,k}^{u,m,t}}\right) \times 100. \tag{12}$$

The store-price 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

[11]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.

index explained by timing of purchases within stores. If the BHI for a consumer is lower than 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.2 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 verify that consumers who purchase relative large quantities pay less within each single cathegory. Further robustness results are relegated to an appendix.

In Table 2, we first show 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, μ_i is an individual fixed effect (FE; we show results with and without this), $\gamma_{m,t}$ is a market×month FE, and X is a vector of regressors: A dummy for age 65 and older, the logarithm of expenditure (our proxy for labor income), the number of shopping trips, and the monthly average of different stores visited. Aguiar and Hurst (2007) use this type of regression, although they do not include the average number of different stores visited.

The left panel of Table 2 shows results for regressions without individual fixed effects. The results in column (1), when only expenditure and age (besides FEs) are included, confirm previous results that consumers 65 and older find lower prices and that higher-spending consumers pay relatively more. This is consistent with the model's interpretation that older individuals' have more time to search while high-wage workers elect to search less.

In columns (2)-(4), we include the average number of (different) stores visited in a month and the number of shopping trips as direct measures of search effort. 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 an identical consumption basket. The inclusion of the average 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 average number of stores visited, gives a highly significant coefficient for trips with a lower R-square of 0.02. However, including the number of shopping trips together with the average 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 average number of stores visited rather than the number of trips that mostly correlates with lower prices although, according to column (4), one more trip (controlling for number of stores visited) lowers the average price paid by 0.05 percent.[12]

In the right panel of Table 2, 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, 2011). Expenditure and age are still significant. The coefficient on age is now identified only from consumers who turn 65 during the sample period, consistent with a clear effect of retirement on time available for shopping. 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 in many stores and these consumers may be more informed obtaining bigger savings from multi-store shopping. The number of trips and the number of stores

[12] This is consistent with the findings of Kaplan et al. (2016) that some stores are cheaper for some goods but not for others.

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 relation 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 versus average UPC-level prices. We estimate the regression

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

where $\mu_{i,c}$ denotes individual $\times$ category FEs, and $\mathrm{QI}^{c}_{i,t}$ is the quantity of category c that individual i consumes relative to the average person in consumer panel. The data form a panel indexed by individual $\times$ category $\times$ time, and the coefficient β captures whether consumers find lower prices for the goods in the categories from which they buy relatively more.

The results presented in Table 3 show that consumers do indeed pay less for goods of which they purchase relatively larger quantities, which is consistent with the rational allocation of time across goods categories. The coefficient to the quantity index of 2.27 in column (1) implies that a one standard deviation increase in the quantity index (about 0.74) results in about 2 percent savings in the category. The coefficient is estimated with very high precision, with a t-statistic greater than 100. Age is not significant, implying that the effect of retirement is captured by the other regressors, for example the category$\times$ individual FEs that cannot be controlled for in the previous regressions.

Log-expenditure is significant with an elasticity of –1.74 while the number of stores visited is highly significant, as is the number of total trips, although the number of trips has less of a price effect than the number of stores visited.

Columns (2) repeats the estimation for the pre-2008, giving similar results. This column is included because the store-price bargain-hunting index can only be calculated for that sample. In column (3), we show the results using this index by running the regression

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

which is similar to the of the previous columns except for the left-hand side being the store-price bargain-hunting index by category. The results obtained for the store-pricing 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. The difference represents gains from the timing of purchases. The estimated coefficients for the store-pricing BHI, in column (3), are smaller than those of the previous columns across the board. In particular, the coefficient to the quantity index drops from 2.28 to 0.53, although it is still extremely significant statistically. The interpretation is that a large fraction of the savings obtained from shopping result from choosing the time to shop in a given store, rather than systematically shopping at stores with relatively lower prices (by category).[13]

[13] The results regarding the relative importance of timing is sensitive to the length of the period used. If, for example, store-prices are averaged over only a week, the timing becomes less important than the store visited. The interpretation of this pattern is that some consumers are able to time their purchases over intervals longer than a week.

The regression results reported in Table 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 goods than for other goods, with almost no variation in prices paid for identical beer UPCs and little variation for cigarettes (followed by milk and sugar substitutes). Laundry detergent displays the largest variation, followed by hot dogs and mayonnaise. 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.

To test whether our results are robust across categories, we estimate the regression

$$\mathrm{BHI}^c_{i,t} = \mu^c_i + \gamma^c_{m,t} + \beta^c \mathrm{QI}^c_{i,t} + X_{i,t}\alpha^c + \epsilon_{i,t}$$

separately for each category c. The data in each regression form an individual $\times$ time panel, and all coefficients, including dummies, can take different values for the different categories.

Table 4 shows the regressions category-by-category. 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 every single category. The sizes of the coefficients to quantity vary (even though the quantity index in these regressions has been standardized to have a mean 0 and a standard deviation of 1 by category for an easier comparison of the coefficients across the 30 regressions). For some categories, including beer, the coefficient is small, which reflects that there is little variation in the prices of those products (for the exact same UPC, that is). In fact, all the categories with a coefficient to quantity that is less than unity are among the categories with the lowest variation in prices paid. The largest coefficient is for photography, one of the categories with the highest price variation. So the variation in coefficients is intuitive, but because our focus is on the qualitative result, we do not model this variation in more detail.

5 Discussion

We compare prices at the UPC level, so our findings provide a lower bound on the savings obtained, because consumers may save by changing to less expensive brands or buying in bulk. However, buying different brands or package sizes to obtain savings brings a potential loss in utility that can be evaluated only by using functional forms, which we avoid in this paper.

If a consumer changes consumption within a category to more expensive brands, this will appear as an increase in the quantity index. The bargain hunting index compares the price paid to the average price of the 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 extend that this does not happen, it will play the role of measurement error and bias our coefficients towards zero.

Our focus is not necessarily on permanent preferences for goods, as most of our regressions include consumer fixed effects which absorb permanent differences. The interpretation is therefore that if a consumer, say, buys a lot of beer in a month where he or she consumes a lot of beer, for example, due to a celebration, the consumer will 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 results 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 examine this issue in two ways, using lagged quantities and using a longer time horizon.

Lagged and current (category) relative quantities are correlated (with a coefficient around 0.45). Both current-month quantities and lagged quantities, see Table A.3, significantly predict current prices. Sales are unlikely to extend across months and we interpret this to imply that preferences are correlated across months and this pattern therefore indicates causality from preferences to prices. (The coefficient is smaller when we use lagged quantities because preferences may vary month-by-month.) Also, 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 regression in Table A.4 for quarterly frequencies. The results are very similar to those obtained at the monthly frequency which supports our interpretation.

6 Conclusion

We find that, consistent with a model of rational price search, consumers pay lower prices for goods of which they consume more, and they pay more for goods of which they consume less. The empirical results provide robust support for the notion that consumers rationally search for best prices. Our results are consistent with those from models of "shoppers" versus "loyals," or "cool" versus "busy," in that we document significant variation in the prices consumers pay on average. In Table A.2 in the appendix, we illustrate the magnitudes of 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).[14] 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 buys less.

Overall, there is substantial heterogeneity across consumers, as previously documented. Our contribution is to document how rational consumers shop across goods and conduct more bargain hunting for the most desired goods.

[14]If the number of categories is not even, the top group has one more category.

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 (2011, January). Best Prices. NBER Working Paper No. 16680, National Bureau of Economic Research, Inc, Cambridge, MA.

Coibion, O., Y. Gorodnichenko, and G. H. Hong (2015, March). The Cyclicality 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 (2016, January). Relative Price Dispersion: Evidence and Theory. NBER Working Paper No. 21931, National Bureau of Economic Research, Inc, Cambridge, MA.

Nevo, A. and A. Wong (2018, forthcoming). The Elasticity of Substitution Between Time and Market Goods: Evidence from the Great Recession. *International Economic Review*.

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. 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, 65+	190,607	0.62	9.21	–27	32
BHI, age<65	360,831	–0.33	8.44	–27	32
BHI, exp. $<$ median exp.	275,596	0.50	9.78	–27	32
BHI, 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 (9). 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 (9) 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 (11). 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 quarter; (4) # stores visited (weekly average) is the weekly average number of different stores that a consumer visits in a given month.

Table 2. The Bargain Hunting Index for Overall Expenditure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pooled OLS				Panelist Fixed Effects			
Log. Expenditure	–0.62***	–0.81***	–1.08***	–0.91***	–0.55***	–0.66***	–0.80***	–0.78***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)
Old (65+)	0.71***	0.40***	0.34***	0.35***	0.36**	0.30*	0.32*	0.29*
	(0.12)	(0.11)	(0.11)	(0.11)	(0.17)	(0.17)	(0.17)	(0.17)
# stores visited (monthly)		0.77***		0.68***		0.32***		0.22***
		(0.03)		(0.03)		(0.02)		(0.02)
# trips (monthly)			0.19***	0.05***			0.10***	0.06***
			(0.01)	(0.01)			(0.01)	(0.01)
Month-Year × 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: $\text{BHI}_{i,t} = \mu_i + \gamma_{m,t} + X_{i,t}\,\alpha + \epsilon_{i,t}$, where $\text{BHI}_{i,t}$ is the bargain hunting index for individual i in month t, μ_i is an individual fixed effect (FE) ($\mu_i = \mu$ in the first four columns), $\gamma_{m,t}$ is a market×month FE, and X is a vector of regressors: A dummy for age 65 and older, the logarithm of total expenditure, the average 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 3. Rational Inattention. Pooled Regressions

	(1) All Years BHI	(2) Pre-2008 BHI	(3) Pre-2008 Store BHI
Quantity Index	2.27***	2.28***	0.53***
	(0.02)	(0.02)	(0.01)
Log. Expenditure	–1.75***	–1.91***	–0.50***
	(0.03)	(0.03)	(0.02)
Old (65+)	0.17	0.09	0.32
	(0.13)	(0.35)	(0.23)
storecount	0.18***	0.27***	0.15***
	(0.01)	(0.01)	(0.01)
(sum) tripcount	0.05***	0.05***	0.02***
	(0.00)	(0.01)	(0.00)
Month-Year × Market FE	Yes	Yes	Yes
Category × Individual FE	Yes	Yes	Yes
Observations	3728872	2126058	2126058
Adj. R-squared.	0.16	0.16	0.15

Notes: Regression: $\text{BHI}_{i,t}^c = \mu_{i,c} + \gamma_{m,t} + \beta\text{QI}_{i,t}^c + X_{i,t}, \alpha + \epsilon_{it}$, where $\text{BHI}_{i,t}^c$ is the category-specific bargain hunting index for individual i in month t, $\mu_{i,c}$ denotes individual× 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 (11). 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 (3), $\text{BHI}_{i,t}^c$ is replaced by a category-specific store BHI, $\text{BHI}_{i,t}^{c,s}$. Standard errors clustered by individual. *** (**) [*] significant at the 1 (5) [10] percent level.

Table 4. The BHI and the QI by Category. Separate Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Quantity Index	0.16***	1.39***	2.23***	0.35***	1.97***	2.38***	3.33***	0.93***	2.29***	2.30***
	(0.02)	(0.14)	(0.04)	(0.08)	(0.06)	(0.05)	(0.12)	(0.12)	(0.07)	(0.06)
Log. Expenditure	–0.17***	–1.55***	–2.50***	–0.40***	–2.01***	–2.17***	–1.82***	–1.71***	–1.69***	–2.71***
	(0.04)	(0.29)	(0.07)	(0.10)	(0.10)	(0.07)	(0.19)	(0.27)	(0.12)	(0.10)
Old (65+)	–0.09	–0.58	–0.14	0.06	0.15	0.20	–1.24*	–0.33	–0.05	0.76*
	(0.14)	(0.77)	(0.27)	(0.54)	(0.35)	(0.29)	(0.71)	(0.77)	(0.36)	(0.39)
# stores visited (monthly)	0.02	0.06	0.24***	0.13***	0.12***	0.17***	0.42***	0.06	0.22***	0.13***
	(0.01)	(0.11)	(0.03)	(0.04)	(0.04)	(0.03)	(0.08)	(0.11)	(0.05)	(0.04)
# trips (monthly)	–0.00	0.05*	0.04***	0.02*	0.07***	0.07***	0.06**	0.09***	0.06***	0.11***
	(0.01)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.01)
Observations	85387	17091	351243	25336	150937	298853	42228	11488	108841	163080
Adj. R-squared.	0.05	0.10	0.13	0.26	0.16	0.17	0.17	0.14	0.17	0.16

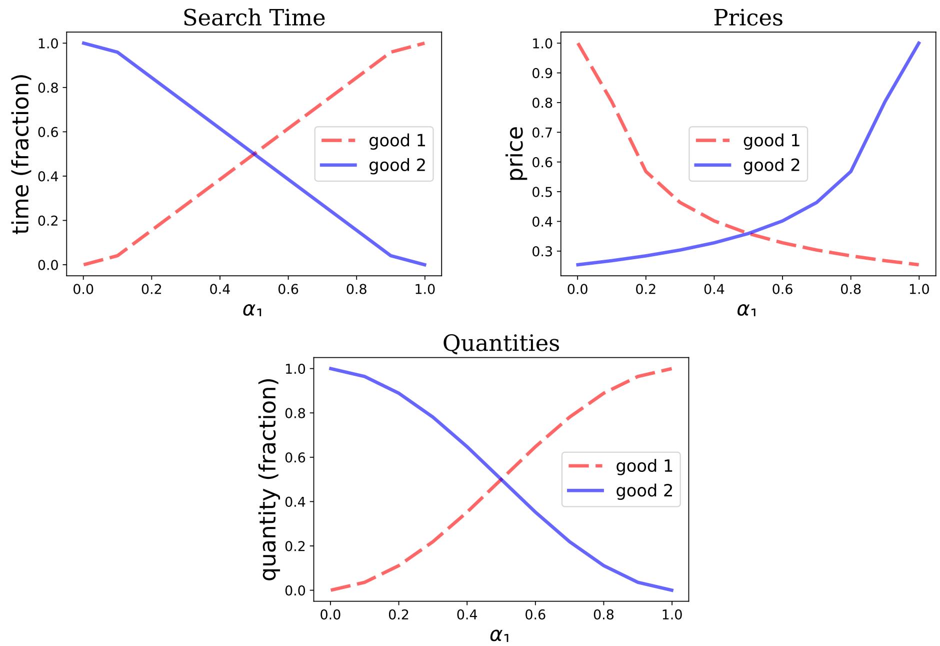
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Quantity Index	2.08***	0.83***	3.74***	2.79***	2.63***	3.43***	1.08***	3.39***	3.06***	2.82***
	(0.06)	(0.06)	(0.09)	(0.08)	(0.07)	(0.07)	(0.03)	(0.08)	(0.08)	(0.08)
Log. Expenditure	–2.45***	–0.48***	–2.80***	–3.79***	–0.99***	–1.44***	–0.60***	–1.09***	–2.28***	–0.96***
	(0.11)	(0.10)	(0.15)	(0.14)	(0.08)	(0.10)	(0.04)	(0.13)	(0.11)	(0.11)
Old (65+)	0.64	0.18	0.41	–0.34	0.05	0.49	0.45**	–0.59	0.01	0.40
	(0.40)	(0.34)	(0.53)	(0.51)	(0.35)	(0.39)	(0.21)	(0.41)	(0.36)	(0.44)
# stores visited (monthly)	0.13***	0.05	0.13**	0.30***	0.14***	0.14***	0.16***	0.12**	0.17***	0.06
	(0.04)	(0.05)	(0.06)	(0.05)	(0.04)	(0.05)	(0.02)	(0.05)	(0.05)	(0.05)
# trips (monthly)	0.07***	0.02	0.10***	0.15***	0.02	0.06***	0.02**	0.04**	0.06***	0.04***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Observations	157666	52248	111656	145363	137280	119403	417846	73326	114039	100484
Adj. R-squared.	0.12	0.13	0.13	0.17	0.17	0.17	0.18	0.17	0.15	0.22

	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
Quantity Index	2.95***	0.62	2.09***	1.95***	2.45***	1.00***	2.56***	2.47***	3.11***	1.66***
	(0.37)	(0.49)	(0.04)	(0.10)	(0.06)	(0.11)	(0.05)	(0.17)	(0.09)	(0.05)
Log. Expenditure	–2.99***	–2.42***	–1.15***	–1.37***	–1.31***	–0.72***	–2.17***	–1.27***	–1.65***	–1.22***
	(0.62)	(0.90)	(0.06)	(0.18)	(0.10)	(0.16)	(0.08)	(0.31)	(0.15)	(0.07)
Old (65+)	–0.10	–0.98	0.40	–0.24	0.11	0.22	0.13	–0.07	–0.64	0.38
	(2.23)	(3.40)	(0.25)	(0.57)	(0.39)	(0.64)	(0.28)	(0.85)	(0.51)	(0.33)
# stores visited (monthly)	0.02	0.71**	0.22***	0.30***	0.18***	0.19**	0.20***	0.10	0.29***	0.16***
	(0.25)	(0.30)	(0.03)	(0.07)	(0.05)	(0.07)	(0.03)	(0.11)	(0.06)	(0.04)
# trips (monthly)	0.14	0.14	–0.00	0.06**	0.06***	0.02	0.06***	0.11***	0.07***	0.05***
	(0.08)	(0.12)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.04)	(0.02)	(0.01)
Observations	4737	2119	341712	40717	132787	20652	197233	20850	67208	217025
Adj. R-squared.	0.16	0.01	0.14	0.12	0.16	0.18	0.18	0.11	0.17	0.13

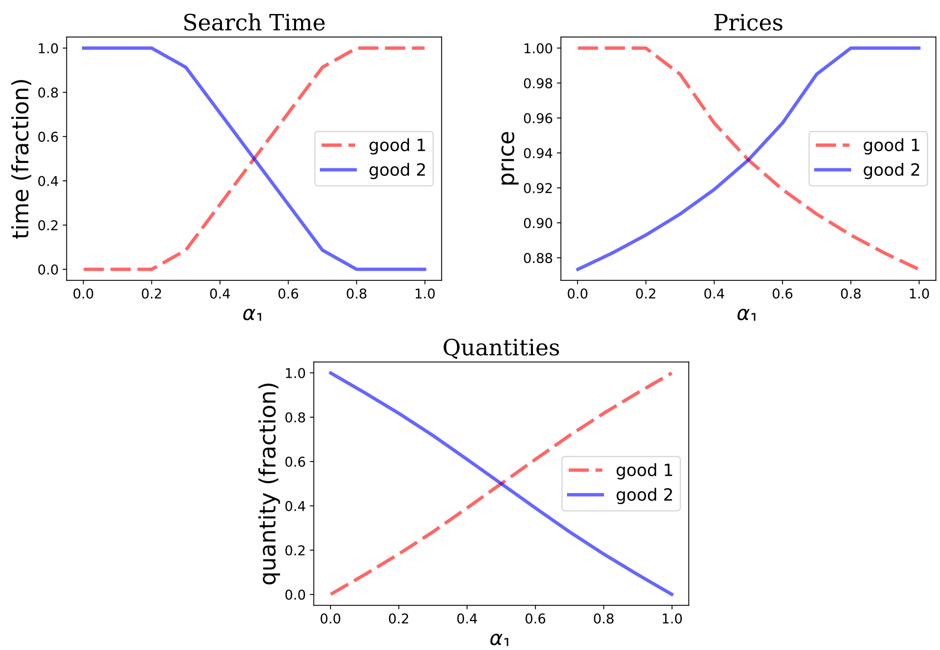
Notes: Regression: $\text{BHI}^c_{i,t} = \mu^c_i + \gamma^c_{m,t} + \beta^c \text{QI}^c_{i,t} + X_{i,t}\alpha^c + \epsilon_{i,t}$, estimated category by category. The quantity index, $QI^c_{i,t}$, is standardized by category (mean 0, sd 1) for easier interpretation. All regressions include market×month FE and individual FE. 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. Standard errors clustered by panelist. *** (**) [*] significant at the 1 (5) [10] percent level.

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

Panel A: High Search Efficiency



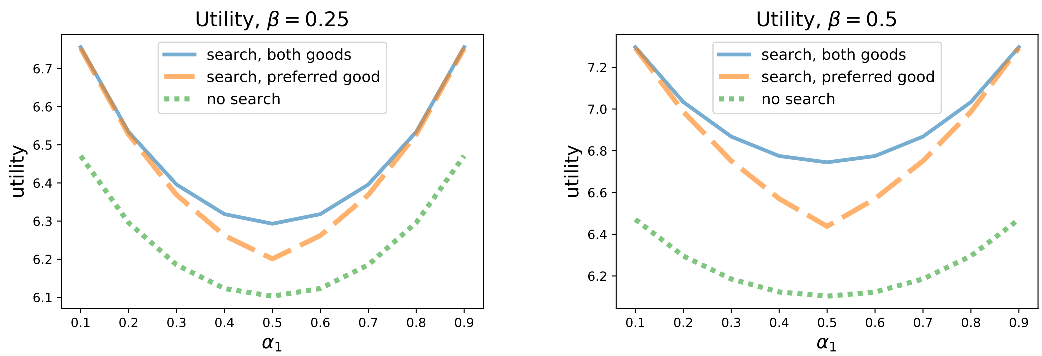
Panel B: Low Search Efficiency



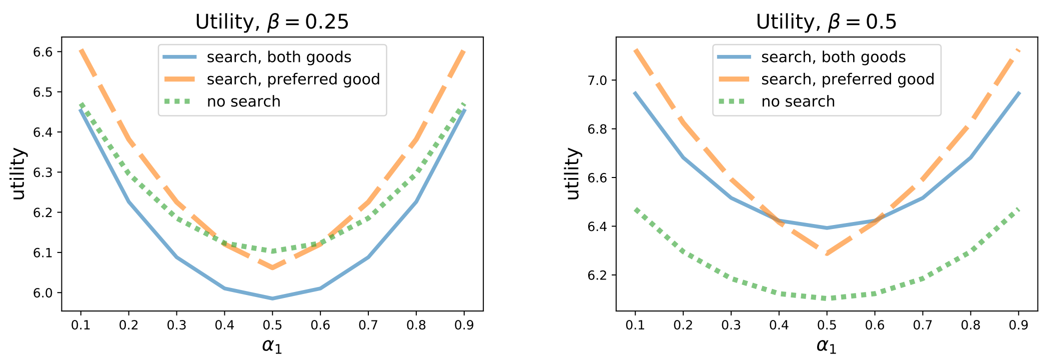
Notes: The figure depicts optimal shopping times, prices and quantities according to the model of Section 2. The model parameters are as follows. $T = 50,\ P_H = 1, \eta = 1,\ \mu = 0.5,\ T_0 = 0,\ W_0 = 50,\ W_1 = 5$. Search efficiency, β, is 0.1 in Panel A and 0.5 in Panel B. In the plots, search time for good i is reported as fraction of the total shopping time $\frac{T_i}{T_1+T_2}$, and similarly for quantity, $\frac{C_i}{C_1+C_2}$. α_1 measures the relative preference for good 1, as $\alpha_1 + \alpha_2 = 1$.

Figure 2. To Search or Not To Search. Utility under Different Scenarios

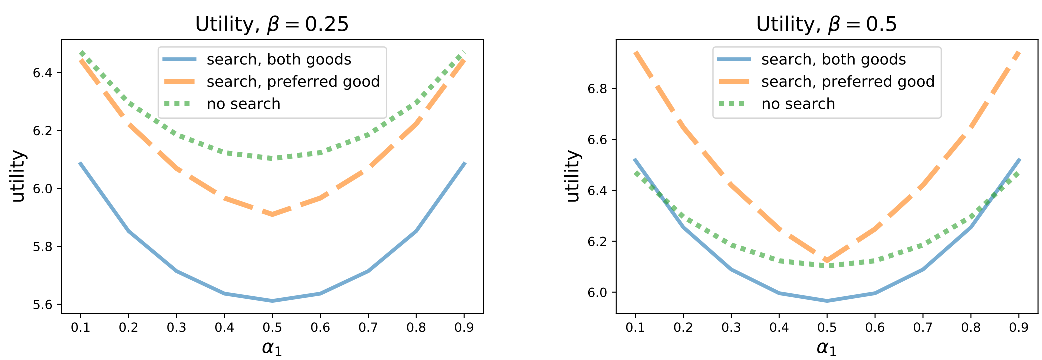
Panel A: no search fixed cost, $T_0 = 0$



Panel B: low search fixed cost, $T_0 = 5$



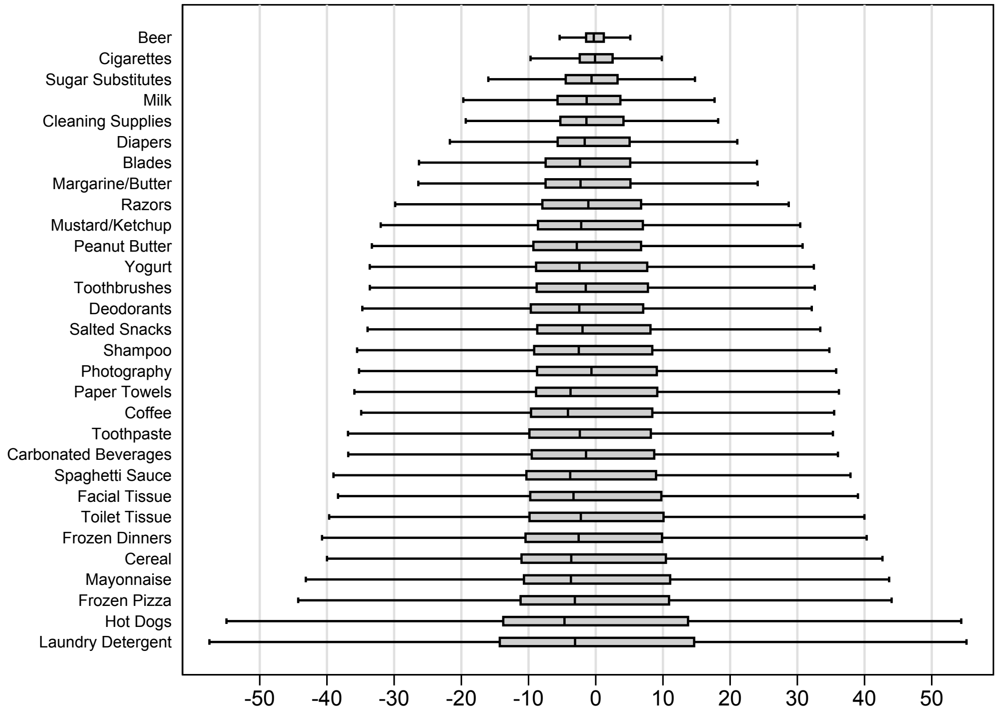
Panel C: high search fixed cost, $T_0 = 10$



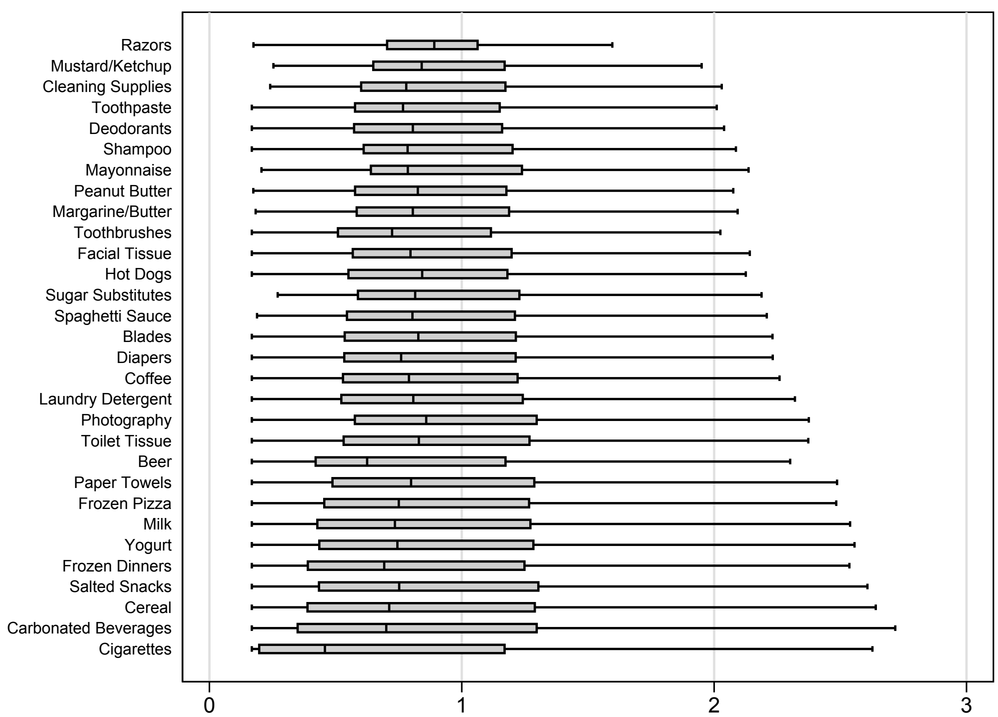
Notes: Each plot depicts utility under three scenarios: no search, search only for the most preferred good (highest α_i), and search for both goods. The model parameters are as follows: $T = 50$, $P_H = 1, \eta = 1$, $\mu = 0.5$, $W_0 = 50$, $W_1 = 5$. Search fixed costs, searching efficiency, and relative preference for good 1 (T_0, β and α_1) vary as indicated in each subplot.

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

Store pricing when consumer search for only some goods.

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 (2018) 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, Menzio, Rudanko, Trachter (2016) 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 of model 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 consumers 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. Average Values of BHI Within Quarters of its Distribution

	(1)	(2)	(3)	(4)	(5)
			BHI in Group		
	All	Q1	Q2	Q3	Q4
Mean BHI	0.00	11.75	1.88	–3.74	–10.21
Mean BHI top-half categories	0.74	13.91	2.69	–3.49	–10.50
Mean BHI bottom-half categories	–2.28	7.17	–0.57	–4.37	–9.60

Notes: The table displays in the first row the overall average value of the bargain hunting index (BHI), normalized to be 0, and the average value for the quarter of consumers with the highest, second-highest, second-lowest, and lowest value of the 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 (we include only consumers who purchase from at least two categories). We then compute two BHIs for top-half and bottom-half categories separately. 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 quarter-group of the overall BHI in columns (2) through (5).

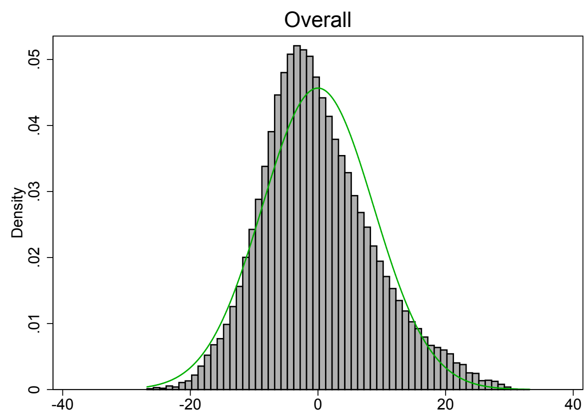
Additional Figures

In Figure A.1, 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.

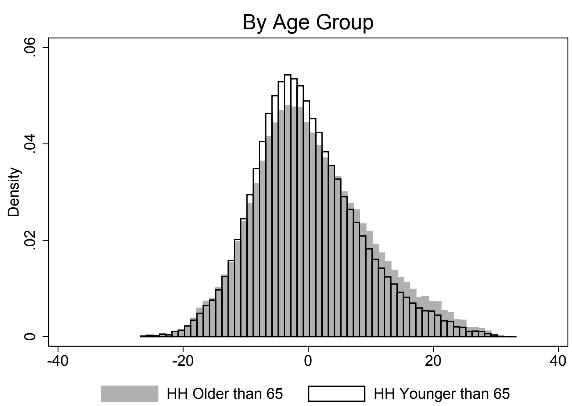
Figure A.2 depicts histograms for the overall BHI and the store BHI. The histograms indicate savings from both store selection (consumers' purchasing products in stores where they are relatively cheaper) and the timing of purchases within a given store.

Figure A.1. The Bargain Hunting Index

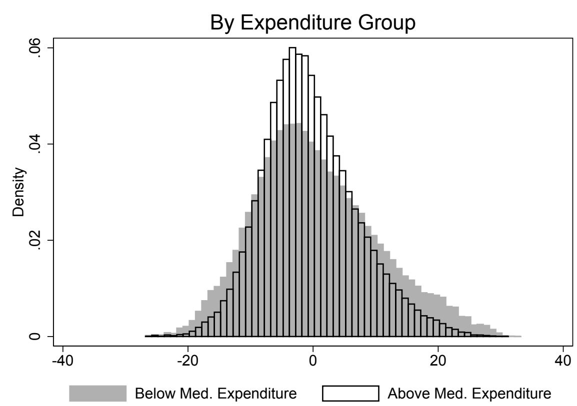
Panel A: Overall



Panel B: By Age Group



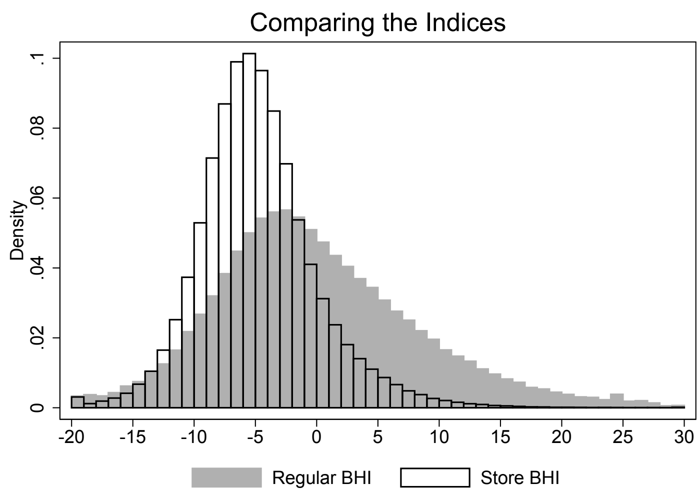
Panel C: By Expenditure Group



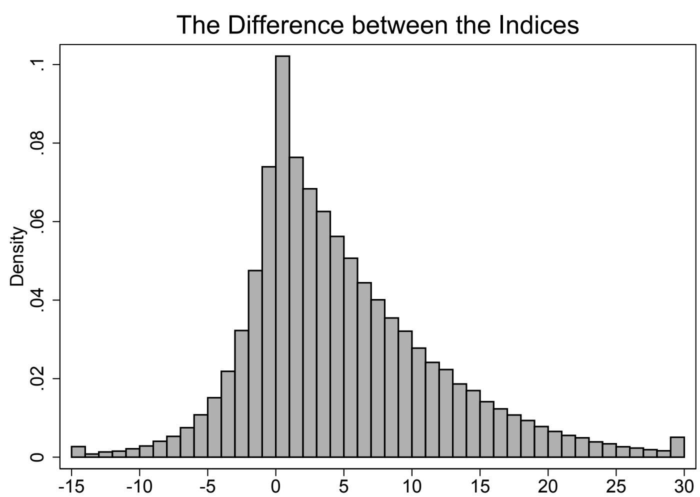
Notes: The BHI index shows how much a consumer saves in percentages compared with the counterfactual of buying his or her consumption bundle at average prices. The BHI index has been normalized to have a mean of 0 every quarter-year by market. *Source:* IRI, all panelist data from 2003 through 2012.

Figure A.2. 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 (9) and represents how much a consumer saves relative to buying at average prices across stores. The store BHI is defined in equation 12 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 2012.

Table A.3. Rational Inattention. Current vs. Lagged Quantity Index

	(1) Current	(2) Lag	(3) Current	(4) Lag
Quantity Index	2.27***	0.80***	2.27***	0.15***
	(0.02)	(0.02)	(0.02)	(0.01)
Log. Expenditure	–1.65***	–0.83***	–1.75***	–0.66***
	(0.05)	(0.05)	(0.03)	(0.03)
Old (65+)	0.08	0.13	0.17	0.16
	(0.10)	(0.11)	(0.13)	(0.14)
# stores visited (monthly)	0.56***	0.57***	0.18***	0.16***
	(0.03)	(0.03)	(0.01)	(0.01)
# trips (monthly)	0.05***	0.05***	0.05***	0.06***
	(0.01)	(0.01)	(0.00)	(0.00)
Month-Year × Market FE	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	No	No
Category × Individual FE	No	No	Yes	Yes
Observations	3754082	3555763	3728872	3539361
Adj. R-squared.	0.03	0.01	0.16	0.14

Notes: Expenditure is aggregated at the monthly level, and store-average prices are calculated at the same frequency. Regression: $\text{BHI}^c_{i,t} = \mu_{i,c} + \gamma_{m,t} + \beta \text{QI}^c_{i,t/t-1} + X_{i,t}, \alpha + \epsilon_{it}$, where $\text{BHI}^c_{i,t}$ is the category-specific bargain hunting index for individual i in quarter t, $\mu_{i,c}$ denotes individual× category fixed effects, , $\gamma_{m,t}$ is a market×month FE, X is a vector of regressors, and $QI^c_{i,t}$ is the quantity index described in equation (11). 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 QI is either current or lagged, as indicated in the column headings. Standard errors clustered by individual. *** (**) [*] significant at the 1 (5) [10] percent level.

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

	(1) All Years BHI	(2) Pre-2008 BHI	(3) Pre-2008 Store BHI
Quantity Index	1.97***	2.00***	0.33***
	(0.02)	(0.03)	(0.01)
Log. Expenditure	–1.61***	–1.92***	–0.35***
	(0.04)	(0.05)	(0.03)
Old (65+)	0.26*	–0.12	0.20
	(0.14)	(0.42)	(0.22)
# trips (quarterly)	0.01*	0.01*	–0.00
	(0.00)	(0.00)	(0.00)
# stores visited (weekly avg.)	0.80***	1.17***	0.52***
	(0.07)	(0.08)	(0.04)
Quarter-Year × Market FE	Yes	Yes	Yes
Category × Individual FE	Yes	Yes	Yes
Observations	1600439	879595	879595
Adj. R-squared.	0.22	0.23	0.22

Notes: Expenditure is aggregated at the quarterly level, and store-average prices are calculated at the same frequency. Regression: $\text{BHI}^c_{i,t} = \mu_{i,c} + \gamma_{m,t} + \beta \text{QI}^c_{i,t} + X_{i,t}, \alpha + \epsilon_{it}$, where $\text{BHI}^c_{i,t}$ is the category-specific bargain hunting index for individual i in quarter t, $\mu_{i,c}$ denotes individual× category fixed effects, , $\gamma_{m,t}$ is a market×quarter FE, X is a vector of regressors, and $QI^c_{i,t}$ is the quantity index described in equation (11). 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 (3), $\text{BHI}^c_{i,t}$ is replaced by a category-specific store BHI, $\text{BHI}^{c,s}_{i,t}$. Standard errors clustered by individual. *** (**) [*] significant at the 1 (5) [10] percent level.