Consumer Search and Price Dispersion in Supermarkets*

Pierre Dubois† Helena Perrone‡

June 2026

Abstract

We study consumer search and demand in French supermarkets using household-level scanner data. We document substantial price dispersion and frequent changes in stores' positions in the price distribution, suggesting that consumers face imperfect information about current prices. We also show that household characteristics associated with time constraints and shopping costs are systematically related to prices paid and store visits. Motivated by this evidence, we estimate a structural model in which consumers search sequentially across stores and choose among differentiated coffee products within each sampled store. The results imply substantial informational frictions: many consumers purchase after observing few stores, weakening substitution across retailers. The full search model implies limited store-level substitution, with a share-weighted own-store elasticity of -0.396 and a share-weighted cross-store elasticity of 0.104. The estimates also imply a choice-utility loss from imperfect price information. Together, these results show that imperfect price information weakens competitive pressure across retailers and lowers the expected utility households obtain from their store choices.

Keywords: imperfect information, price dispersion, search costs, demand estimation, retail markets

JEL Codes: D4, D83, L11, L66

*We thank Aviv Nevo, Larbi Alaoui, Volker Nocke, Robin Lee, Jean Fran çois Houde, Paul Grieco, Michelle Sovinsky, Jose Luis Moraga-Gonzalez, Sandro Shelegia, Matthijs Wildenbeest and participants in several seminars and conferences for comments. Perrone gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224. Dubois acknowledges financial support from the Agence Nationale de la Recherche (ANR) under grant No. ANR17-EURE-0010 (Investissements d'Avenir program) and grant ANR-20-ORAR-0008 PPFM (EST010223 1 EST010924 1).

†Toulouse School of Economics, pierre.dubois@tse-fr.eu

‡Toulouse Business School, helenaperrone@gmail.com

1 Introduction

Traditional demand models often assume that consumers observe all relevant prices before choosing where to buy. In supermarkets, this assumption is unlikely to hold. Prices vary across stores and over time because of temporary promotions, and these promotions are not always known before the shopping trip. Consumers who want to identify the best current price must acquire information, which requires time, travel, and attention.

These informational frictions matter for demand and competition. When consumers observe only a subset of current prices, observed purchases reflect both preferences and information sets. A store may face limited substitution not only because consumers prefer it, but also because many consumers do not observe prices at competing stores before buying. Ignoring this channel can therefore overstate how strongly consumers respond to price differences across retailers.

This paper studies consumer search and retail demand using household-level scanner data from French supermarkets between 1999 and 2001. This period predates widespread online grocery shopping and online price-comparison tools, making it a useful setting for studying price information acquired through physical shopping trips and other offline sources. The data record food purchases by participating households, including the date, price, product characteristics, and retail chain. The reduced-form analysis covers beer, coffee, cola, and whisky. The structural estimation focuses on coffee, a frequently purchased category with substantial price dispersion and broad store availability.

We first document patterns consistent with imperfect price information. Prices exhibit substantial dispersion even for narrowly defined products, in line with broader evidence on retail price dispersion (e.g. Lach, 2002; DellaVigna and Gentzkow, 2019; Hitsch et al., 2021). Stores also move frequently in the price distribution, making it difficult for consumers to know which retailer currently offers the best deal. Finally, household characteristics associated with time constraints and shopping costs are systematically related to prices paid and store visits. Households observed visiting more stores pay lower prices and are more likely to obtain the lowest observed market price, while several characteristics associated with higher search costs are associated with less shopping across stores or higher prices paid.

Motivated by this evidence, we estimate a structural model in which consumers search sequentially across differentiated stores and choose among differentiated coffee products within each sampled store. Conditional on visiting a store, the consumer observes the prices and characteristics of the coffee products available there; the resulting within-store inclusive value enters the store-level utility that governs search. Store preferences may vary with household characteristics, allowing for horizontal differentiation that is especially relevant in physical retail because location, access, and shopping convenience differ across households. The model's sampling probabilities are implemented as empirical store-exposure measures constructed from recorded purchase trips by similar households, excluding the household's own trips. These measures summarize the stores that households of a given type are likely to encounter in their broader grocery-shopping environment. Demand and search behavior are then estimated from purchases in the product category being modeled. The main results are robust to alternative measures of store exposure, including shrinkage toward uniform exposure over available stores.

The coffee estimates show that search frictions change the measurement of demand. Consumers often purchase before sampling many competing stores, so a store's demand depends not only on its price and characteristics but also on whether it enters consumers' information sets during search. This changes the implied substitution patterns. The share-weighted own-store elasticity is -0.396 in the full search model, compared with -1.589 in a benchmark that holds fixed each store's position in the indirect-utility ordering; the corresponding cross-store elasticity is 0.104, compared with 0.397. Thus the direct response to price changes is substantially dampened once price changes are allowed to move stores across the utility ranking that enters consumers' stopping decisions. The same mechanism generates a consumer-side loss: because search is costly, expected indirect utility under the estimated search process is below the indirect utility of the best store alternative available to the consumer.

The paper contributes first to the empirical literature on consumer search and differentiated demand. Existing work shows that limited information affects choices, substitution patterns, demand elasticities, and market power in markets such as online retail, insurance, mortgages, hotels, automobiles, and mutual funds (Hortaçsu and Syverson, 2004; De los Santos et al., 2012; Honka, 2014; Allen et al., 2014; Koulayev, 2014; Moraga-González et al., 2023); see Honka et al.

(2019) and Ursu et al. (2025) for surveys of this literature. Our setting is a brick-and-mortar grocery market in which prices move frequently and stores are horizontally differentiated. In grocery markets, Seiler (2013) estimates search costs from scanner data for a storable product category, focusing on within-store price uncertainty and purchase timing; we instead model search across stores, which is the margin on which retail competition operates. Building on the rank-based identification logic of Hortaçsu and Syverson (2004), we extend the approach to a setting with horizontally differentiated stores and empirically measured, non-uniform store exposure. The household-level data allow us to construct heterogeneous store-exposure probabilities from observed shopping behavior and to let store preferences vary with household characteristics. The search mechanism is therefore estimated conditional on store differentiation, rather than used as a substitute for persistent store heterogeneity. This is important in physical retail markets, where location, access, and shopping convenience make the same retailer more attractive to some households than to others.

Second, the paper relates to work on consideration sets, incomplete availability, and restricted choice in demand estimation. In these models, consumers choose from only a subset of the alternatives available in the market, and ignoring this restriction can bias estimates of preferences and substitution patterns (Conlon and Mortimer, 2013; Abaluck and Adams-Prassl, 2021; Crawford et al., 2021). The restriction may arise from product availability, distribution, awareness, or latent consideration; Abaluck and Adams-Prassl (2021) show how consideration probabilities can be identified from asymmetries in demand responses without observing consideration directly. In our case, the restriction arises from costly search. Before buying, households observe products' prices only at the stores they sample, but they can expand their information set by paying the cost of sampling another store. We therefore model the formation of the information set as part of the consumer's decision problem, rather than treating restricted choice as an exogenous or latent constraint on demand.

Third, the paper connects to work on retail price dispersion and price transparency. Recent studies document substantial pricing heterogeneity across retailers and over time, even in markets with repeated purchases and large chains (Kaplan and Menzio, 2015; DellaVigna and Gentzkow, 2019; Hitsch et al., 2021). In particular, Kaplan et al. (2019) show that most of this

dispersion is *relative*: the identity of the low-price store for a given product varies across products and over time, a pattern our quartile-transition evidence confirms for French supermarkets and which sustains the value of search. In related gasoline markets, changes in price information affect both consumer search and firms' pricing incentives (Chandra and Tappata, 2011; Pennerstorfer et al., 2020; Martin, 2024). We show how imperfect price information changes estimates of grocery purchase substitution patterns. Product-level prices affect both within-store product choice and the value of continuing to search across stores, so estimated elasticities depend jointly on preferences, the store-exposure process, and search costs.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 documents price dispersion and household shopping behavior. Section 4 develops the consumer search framework. Section 5 presents the structural estimates, price elasticities, and a decomposition of the utility cost of imperfect price information. Section 6 concludes.

2 Data and Product Choice

The data come from a representative household scanner panel covering all regions of France from 1999 to 2001. Participating households record each food purchase with a scanner. For each transaction, we observe the date, price, quantity, product characteristics, and the retail chain where the purchase was made. The panel also contains household demographics.

The data identify the retail chain but not the exact physical outlet within a chain. We therefore define a store as a retail chain in a region. Throughout the paper, the term store refers to this chain-region alternative. Similarly, a store visit means a store choice observed on a recorded purchase trip: if a household buys something at a chain in a region on a given date, we observe a visit to that store. The data do not record trips that generated no purchase. This distinction is important for the interpretation of store visits in the reduced-form evidence and for the construction of empirical store-exposure measures used in the structural model.

We analyze four product categories: beer, coffee, cola, and whisky. Within each category, products are defined using detailed scanner characteristics such as brand, package size, packaging, and other observable attributes. Throughout the paper, a product refers to these narrowly defined items, which closely approximate SKUs. The structural model later focuses on search

across stores for products within a category, rather than search over the category as a whole. We choose these categories because search incentives may vary with purchase frequency, storability, and expenditure per purchase occasion. Purchases in coffee, beer, and cola are relatively frequent, while whisky is purchased less often.

Household characteristics used in the analysis include age, household size, education, income, employment status, whether the household lives in a rural area, and the presence of young children. We also construct variables related to shopping behavior and storage capacity, including the number of recorded store visits per week, purchase frequency, car ownership, pantry availability, and dog ownership as proxies for inventory-holding capacity and housing size.

The sample period is useful for studying supermarket search because grocery shopping was conducted almost exclusively in brick-and-mortar stores. Online grocery retail and online price comparison were negligible in France at the time, so consumers who wanted to compare supermarket prices generally had to acquire information through physical shopping trips or other offline sources.

Table 1 reports summary statistics for the estimation sample. The household panel exhibits substantial heterogeneity in demographics and shopping environment, including household size, income, rural location, and car ownership. Conditional on a household-week with at least one recorded purchase, households are observed buying from 1.57 stores on average, indicating that multi-store shopping is common but limited in scope within a week. Purchase frequency differs substantially across categories: coffee is observed much more frequently than whisky, consistent with differences in purchase frequency across staple and occasional categories.

3 Price Dispersion and Shopping Behavior

We begin by documenting empirical patterns that motivate a search-based interpretation of supermarket purchases. First, prices differ substantially across stores even for narrowly defined branded products. Second, stores move frequently in the price distribution over time, so the identity of the low-price store is not stable. Third, household shopping behavior is related to transaction prices in ways that are consistent with costly price information. These facts do

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Household characteristics					
Age of household head (years)	45.41	14.96	18	94	11,572
Baby in household	0.195	0.396	0	1	11,572
Inactive household head	0.396	0.489	0	1	11,572
Number of cars	1.371	0.728	0	7	11,572
Pantry capacity	1.313	1.104	0	3	11,572
Number of dogs	0.438	0.756	0	8	11,572
Rural household	0.513	0.500	0	1	11,572
High-income household	0.308	0.462	0	1	11,572
Household size	2.976	1.436	1	9	11,572
Education level of household head	2.469	1.353	0	5	8,976
Shopping and purchase characteristics					
Recorded store visits per household-week	1.571	0.783	1	10	531,687
Number of category purchases over sample period	15.00	21.33	1	154	35,434
Quantity per purchase					
Beer (liters)	3.255	2.695	0.500	54	45,023
Coffee (kg)	0.473	0.301	0.250	6	122,362
Cola (liters)	3.367	3.067	0.330	72	86,127
Whisky (liters)	0.785	0.236	0.700	5	7,642

Notes: The table reports summary statistics for households and recorded shopping behavior in the sample used in the empirical analysis. Baby in household, inactive household head, rural household, and high-income household are indicator variables equal to one when the condition holds and zero otherwise. Education level is coded as a categorical variables. Recorded store visits are measured at the household-week level. The number of category purchases is measured over the full sample period. Quantities per purchase are reported separately for each product category.

not by themselves identify search costs, but they provide the empirical basis for the structural model developed in the next section.

3.1 Price Dispersion Across Retailers and Time

To study price dispersion, we focus on two narrowly defined products within each category. Products are defined by detailed observable characteristics such as brand, package size, packaging, and variety. Conditional on these characteristics, remaining price variation primarily reflects differences across stores and time rather than differences in the physical product.

Transactions of physically identical products occur at different stores and in different weeks. From the consumer's perspective, these transactions may therefore differ because stores offer different locations, shopping environments, services, and promotional activity. For this reason, part of the raw price dispersion may reflect store differentiation rather than pure price differences.

Table 2 reports descriptive statistics for transaction prices. Substantial price dispersion is present in all categories. Depending on the product, prices at the upper end of the distribution exceed prices at the lower end by economically meaningful amounts. For example, the ratio between the 95^{th} and 5^{th} percentiles reaches 1.81 for one cola product and 1.70 for one coffee product, indicating large differences in prices paid for tightly defined goods.

To account for persistent differences across stores and time periods, we also compute residual prices from regressions including chain, region, month, and year fixed effects, as well as store-type controls. The residual price measures therefore capture within-product price variation net of observable retailer and temporal heterogeneity.

The residual price statistics show that substantial dispersion remains after controlling for retailer and time effects. For cola, the interquartile difference in residual log prices reaches 0.21 and the Q95–Q5 difference reaches 0.48. Residual dispersion is also economically meaningful for coffee, where the Q95–Q5 difference reaches 0.45, and for beer, where it reaches 0.23. Whisky displays more modest residual dispersion. These patterns suggest that consumers face meaningful variation in prices across stores and periods that cannot be attributed only to persistent retailer characteristics.

Table 2: Descriptive Statistics of Pricing Patterns

		Uncontrolled Price Dispersion					Residual Log Price Dispersion	
Category	Product	Avg Price	Standard Deviation	Coef of Variation	Q75/Q25	Q95/Q5	Q75-Q25	Q95-Q5
Beer	A	0.98	0.10	0.10	1.21	1.27	0.12	0.23
	B	1.30	0.10	0.07	1.04	1.26	0.03	0.14
Cola	A	0.71	0.05	0.07	1.08	1.18	0.21	0.48
	B	0.49	0.11	0.23	1.28	1.81	0.21	0.48
Coffee	A	8.73	0.64	0.07	1.10	1.22	0.06	0.16
	B	4.49	0.77	0.17	1.34	1.70	0.29	0.45
Whisky	A	15.5	1.65	0.11	1.01	1.44	0.06	0.21
	B	14.9	0.63	0.04	1.02	1.05	0.02	0.06

Notes: The table reports price dispersion for selected products in each product category. The left panel reports unconditional transaction-price dispersion: mean price, standard deviation, coefficient of variation, and percentile ratios. The right panel reports dispersion in residual log prices. Residual log prices are obtained from regressions of log prices on chain, region, month, and year fixed effects, as well as store-type controls. Residual dispersion is measured as 75th–25th and 95th–5th percentile differences.

3.2 Dynamics of Relative Prices

If the identity of the lowest-price store were stable over time, consumers could gradually learn where to shop without repeatedly acquiring price information. In contrast, frequent changes in stores' relative price positions make past price information less useful and increase the value of current search.

To study the dynamics of relative prices, we assign stores in each week to quartiles of the cross-sectional price distribution for a given product. Stores in quartile 1 are in the lowest-price quartile in that week, while stores in quartile 4 are in the highest-price quartile. We then compute empirical transition probabilities across quartiles from one week to the next. Table 3 reports the resulting transition matrices.

The transition matrices reveal substantial mobility across price quartiles. Although there is some persistence along the diagonal, off-diagonal transition probabilities are large in all product categories. Stores that are among the lowest-price options in one week often move to higher price quartiles in the following week, while stores that are initially expensive often become relatively cheaper. These patterns are consistent with temporary promotions and rotating pricing strategies, and with the relative price dispersion documented by Kaplan et al. (2019).

The degree of persistence also varies across products. For example, Cola A displays relatively persistent rankings, with stores in the lowest-price quartile remaining there with probability

0.641. By contrast, for several other products transition probabilities are more dispersed. For Beer A, a store initially in the lowest-price quartile remains there with probability only 0.472, so less than half of these stores remain in the lowest-price quartile one week later. Similar patterns arise for whisky and coffee products.

Overall, stores frequently change their position in the price distribution. Because no store persistently occupies the low-price end of the distribution, consumers cannot rely solely on past experience to identify the cheapest options. Information about current prices and promotions must therefore be updated frequently, providing a natural motivation for a model with costly consumer search.

3.3 Shopping Behavior and Transaction Prices

The final set of descriptive evidence links household shopping behavior to the prices paid. If consumers face costs of acquiring price information or visiting multiple stores, households with higher opportunity costs of time or lower mobility should shop less broadly and pay higher prices on average. Conversely, households observed purchasing at more stores should be more likely to encounter low prices and temporary promotions. These regressions are not intended to identify a particular search protocol. Rather, they document whether the data display the basic correlations one would expect when price information is costly.

We construct category-level price indices that measure how much more or less than the market average a household pays for a given basket of goods. We also examine the probability that a household purchases at the lowest observed market price in a given period. The specifications include household characteristics associated with the opportunity cost of time, as well as store and time fixed effects.

The opportunity cost of time is proxied using household demographics and labor-market status, including age, income, employment status, education, household composition, and the presence of young children. We additionally control for factors related to shopping and inventory behavior, such as purchase frequency and car ownership.

We also include the number of recorded store visits made by the household each week, constructed from purchases across all food categories. This variable counts the distinct stores

Table 3: Store Transition Probabilities Across Price Quartiles

		Beer A				Beer B			
		Rank at $t+1$				Rank at $t+1$			
		1	2	3	4	1	2	3	4
Rank at t	1	0.472	0.204	0.163	0.120	0.484	0.230	0.138	0.103
	2	0.288	0.293	0.234	0.142	0.332	0.361	0.214	0.077
	3	0.195	0.214	0.320	0.235	0.150	0.199	0.423	0.198
	4	0.189	0.118	0.281	0.341	0.129	0.097	0.200	0.523
	Obs	2,642				4,489			
		Coffee A				Coffee B			
		Rank at $t+1$				Rank at $t+1$			
		1	2	3	4	1	2	3	4
Rank at t	1	0.583	0.071	0.161	0.146	0.408	0.198	0.200	0.157
	2	0.415	0.171	0.220	0.111	0.335	0.264	0.255	0.134
	3	0.403	0.098	0.273	0.201	0.235	0.265	0.287	0.191
	4	0.351	0.059	0.208	0.337	0.192	0.240	0.304	0.192
	Obs	4,558				1,392			
		Cola A				Cola B			
		Rank at $t+1$				Rank at $t+1$			
		1	2	3	4	1	2	3	4
Rank at t	1	0.641	0.143	0.119	0.080	0.497	0.197	0.161	0.067
	2	0.264	0.467	0.132	0.113	0.353	0.242	0.170	0.104
	3	0.235	0.193	0.375	0.184	0.303	0.250	0.230	0.143
	4	0.121	0.106	0.164	0.587	0.278	0.185	0.166	0.232
	Obs	12,740				1,237			
		Whisky A				Whisky B			
		Rank at $t+1$				Rank at $t+1$			
		1	2	3	4	1	2	3	4
Rank at t	1	0.438	0.200	0.182	0.117	0.522	0.159	0.147	0.083
	2	0.286	0.289	0.231	0.131	0.408	0.185	0.223	0.076
	3	0.233	0.185	0.313	0.202	0.202	0.174	0.261	0.206
	4	0.181	0.177	0.247	0.301	0.142	0.162	0.311	0.284
	Obs	1,586				862			

Notes: The table reports empirical one-week transition probabilities across quartiles of the cross-sectional price distribution for each product. Rank 1 corresponds to the lowest-price quartile and rank 4 to the highest-price quartile. Rows indicate a store's price rank in week t; columns indicate its price rank in week $t+1$. The "Obs" row reports the number of store-week observations used to compute each product-specific matrix.

at which the household is observed purchasing during the week. It does not necessarily measure all physical shopping trips, because trips with no purchase are not recorded. Nevertheless, recorded visits provide an observable proxy for the breadth of the household's shopping activity and exposure to prices.[1]

Table 4 reports the relationship between shopping behavior and transaction prices. Households with more recorded store visits pay significantly lower prices and are more likely to purchase at the lowest observed market price. The estimated effects are economically meaningful and consistent across specifications.

Several proxies for the opportunity cost of time are also systematically related to prices. Households with young children pay higher prices on average and are less likely to obtain the lowest available price, consistent with time constraints limiting search. By contrast, inactive households pay lower prices and are more likely to purchase at the lowest observed price. Car ownership is also associated with lower transaction prices, potentially reflecting lower transportation costs of visiting multiple retailers.

Table 5 examines the determinants of recorded store visits directly. Households with lower opportunity costs of time tend to be observed purchasing at more stores and are more likely to shop at multiple chain-region alternatives during a given week. Inactive households and larger households exhibit significantly higher shopping intensity, whereas richer, rural, and more educated households visit fewer stores on average.

Overall, the reduced-form evidence is consistent with the presence of economically meaningful search frictions. Households that appear to shop more broadly systematically obtain lower prices, while households facing higher opportunity costs of time search less intensively and pay more on average. These findings motivate the structural search model developed in the next section.

[1]The OLS estimates are subject to the concern that search intensity and prices are jointly determined. The structural model in Section 4 provides a framework in which search intensity and prices are jointly determined by primitive preferences and search costs, and interprets the patterns in Tables 4 and 5 as equilibrium outcomes of that decision problem.

Table 4: Opportunity Cost of Time and Transaction Prices

	Transaction Price Index		Paying Lowest Price	
	(1)	(2)	(3)	(4)
	OLS	OLS	Probit	Probit
Recorded store visits	-0.103***	-0.099***	0.032***	0.031***
	(0.014)	(0.014)	(0.002)	(0.002)
Age		-0.023***		0.010***
		(0.006)		(0.001)
Age squared		0.000***		-0.000***
		(0.000)		(0.000)
Baby in household		0.104***		-0.018***
		(0.035)		(0.007)
Inactive household head		-0.049*		0.033***
		(0.025)		(0.005)
Purchase frequency	0.001***	0.002***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Number of cars	-0.382***	-0.375***	0.054***	0.052***
	(0.064)	(0.064)	(0.013)	(0.013)
Pantry capacity	-0.002	0.008	0.037***	0.034***
	(0.023)	(0.023)	(0.004)	(0.004)
Number of dogs	-0.050**	-0.042*	0.014***	0.012***
	(0.022)	(0.022)	(0.004)	(0.004)
Rural household	-0.030	-0.035	-0.020	-0.018***
	(0.022)	(0.022)	(0.004)	(0.004)
High-income household	-0.042*	-0.027	0.007	0.006
	(0.023)	(0.023)	(0.004)	(0.005)
Household size controls	Yes	Yes	Yes	Yes
Education controls	Yes	Yes	Yes	Yes
Store fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	445,038	445,038	415,438	415,438

Notes: The table reports the relationship between recorded shopping activity and prices paid. Columns (1)–(2) are OLS regressions in which the dependent variable is the household's transaction price index. Columns (3)–(4) report average marginal effects from Probit regressions in which the dependent variable is an indicator equal to one if the household pays the lowest observed market price in the relevant category-period. Recorded store visits count the number of distinct stores at which the household is observed purchasing during the week, constructed from purchases across all food categories. All specifications include household size and education controls, store fixed effects, and time fixed effects as indicated. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Determinants of Store Visits

	Store Visits OLS		Store Visits Poisson		Store Visits > 1 Probit	
Age	0.018***	(0.000)	0.011***	(0.000)	0.024***	(0.001)
Age squared	-0.000***	(0.000)	-0.000***	(0.000)	-0.000***	(0.000)
Baby in household	-0.008*	(0.004)	-0.004	(0.003)	0.010	(0.007)
Inactive household head	0.069***	(0.003)	0.043***	(0.002)	0.120***	(0.005)
Number of cars	-0.004	(0.008)	-0.002	(0.005)	0.007	(0.013)
Pantry capacity	-0.005	(0.003)	-0.003	(0.002)	-0.021***	(0.005)
Number of dogs	-0.007**	(0.003)	-0.004**	(0.002)	-0.012**	(0.004)
Rural household	-0.082***	(0.002)	-0.052***	(0.002)	-0.107***	(0.004)
High-income household	-0.022***	(0.003)	-0.014***	(0.002)	-0.019***	(0.005)
Household Size	Yes	Yes	Yes	Yes		
Education Level	Yes	Yes	Yes	Yes		
Store Fixed Effects	Yes	Yes	Yes	Yes		
Time Fixed Effects	Yes	Yes	Yes	Yes		
N	338,961		338,961		338,961	

Notes: The table reports regressions of household shopping intensity on household characteristics. The dependent variable in the OLS and Poisson specifications is the number of distinct stores at which the household is observed purchasing during the week. The dependent variable in the Probit specification is an indicator equal to one if the household is observed purchasing at more than one store during the week. The Probit column reports average marginal effects. Store visits are constructed from purchases across all food categories and do not include trips with no recorded purchase. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 Consumer Search Framework

This section develops the search framework used in the empirical analysis. A store is defined as a chain-region alternative, as described in the data section. Before making a purchase in a product category, consumers decide which stores to sample. For each sampled store, the consumer observes the prices and characteristics of the narrowly defined products available in that category. The consumer then chooses the variant that maximizes indirect utility within the category. The store-level indirect utility that enters the search problem is therefore the indirect utility from purchasing the consumer's preferred variant in that category at that store.

The model builds on Hortaçsu and Syverson (2004), but adapts their framework to grocery markets in two ways. First, stores are not ordered by price alone. Store utility depends on the full set of product prices and characteristics in the category, chain-region attributes, period-specific unobserved demand shocks, and household characteristics. Second, the empirical store-exposure probabilities that implement the model's sampling probabilities are constructed from recorded purchase trips rather than imposed to be uniform. We interpret them as empirical measures of store exposure: the probability that a household type encounters each store alternative through its broader grocery-shopping environment.

Households choose how intensively to search across stores before making a category purchase. Search is costly because adding another store to the consideration set requires time, travel, planning, and attention before the full set of category prices is observed. Conditional on sampling a store, however, the consumer observes the relevant product prices and characteristics without further search costs. Weekly price changes and promotions make the identity of the best store uncertain, so the model formalizes the trade-off between the expected gain from sampling an additional store and the cost of doing so.

4.1 Model of Consumer Behavior

Before making a category purchase, consumers sequentially sample stores. Search costs are heterogeneous across consumers, and the cost of the first store draw is normalized to zero. Products are narrowly defined items inside a category. Consumers compare stores, while product-level prices and characteristics within the category enter store utility through the value of the preferred product available at that store.

Let i index household type and let j denote a store. Consumers of type i know the distribution of the indirect utility u_{ijt} obtained from visiting store j in period t and buying the preferred product variant in that category in that store. At the product level, let B_{jt} denote the set of product variants available in store j and period t. The utility of product $b \in B_{jt}$ is

$$w_{ijbt} = x_{jbt}\beta - \alpha p_{jbt} + \varepsilon_{ijbt} \tag{1}$$

where x_{jbt} are observed product characteristics, p_{jbt} is the product-level price, α is the marginal disutility of income, and ε_{ijbt} is an i.i.d. type I extreme value product taste shock.

Within store j, the consumer chooses the product with highest indirect utility within that category:

$$b^*_{ijt} = \arg\max_b w_{ijbt}$$

The expected within-store value (the store-level indirect utility) is the logsum over the included

product variants, up to a store-level unobservable v_{ijt} and a common component δ_{ijt}:

$$u_{ijt} \equiv \delta_{ijt} + \underbrace{\log\left[\sum\nolimits_{b \in B_{jt}} \exp\left(x_{jbt}\beta - \alpha p_{jbt}\right)\right]}_{=E[w_{ijb^*_{ijt}t}]} + v_{ijt}. \tag{2}$$

using the usual formula (Small and Rosen, 1981), with the Euler constant absorbed into the normalization.

The common component δ_{ijt} captures observed and unobserved determinants of store utility that are not summarized by product prices and characteristics. It may vary across household types and stores, allowing the same store to have different value for different consumers because of location, access, shopping convenience, and other non-price attributes. The empirical specification below maps this component to observed controls and fixed effects.

Within each household type, consumers share the same distribution of store utilities and sampling probabilities, but search costs vary across consumers. During a purchase occasion, consumers draw stores sequentially. After each draw, the sampling probabilities are renormalized over the stores that have not yet been inspected. Perfect recall implies that the best previously inspected store remains available for purchase.

Let $F_{it}(\cdot)$ denote the belief distribution of indirect utilities u_{ijt} for a type-i consumer. A consumer with search cost c_i who has already found the highest indirect utility among sampled stores, denoted u^*_{it}, searches once more if the cost of an additional draw is no larger than the expected gain from finding a better alternative:

$$c_i \leqslant \int_{u^*_{it}}^{\bar{u}_{it}} \left(u - u^*_{it}\right) dF_{it}\left(u\right) \tag{3}$$

where $\bar{u}_{it}$ is the upper bound of the support of $F_{it}\left(\cdot\right)$.

Consumers know $F_{it}(\cdot)$ and the support of indirect utilities. They can therefore evaluate the position of any realized store utility in the market utility distribution.

We order these N_i store utilities from lowest to highest:[2]

$$u_{i1t} < u_{i2t} < \cdots < u_{iN_it}.$$

For simplicity, we assume that there are no ties in store utility.[3]

As all indirect utilities of stores are strictly different, we obtain:

$$F_{it}(u) \equiv P(u_{ijt} \leq u) = \sum\nolimits_{k=1}^{N_i} \phi_{ikt} \mathbf{1}_{\{u_{ikt} \leq u\}} \tag{4}$$

where ϕ_{ikt} is the first-draw probability that the store ranked k is sampled by a consumer of type i in period t (this probability is known by consumers and common to all consumers of type i in the market) and $\mathbf{1}_{\{u_{ikt} \leq u\}} = 1$ if and only if $u_{ikt} \leq u$.

Using (3) and (4) yields the following cutoff points for the search cost distribution:

$$c_{ij}^t \equiv \sum\nolimits_{k=j+1}^{N_i} \phi_{ikt} (u_{ikt} - u_{ijt}) \tag{5}$$

where c_{ij}^t is the search cost level that makes a consumer of type i whose current best option is store j indifferent between stopping and searching once more. Because store j has already been sampled, stores with indirect utility lower than u_{ijt} do not contribute to the expected gain from an additional search. The cutoff c_{ij}^t is therefore the expected incremental value of drawing one more store, conditional on the consumer's current best option. Consumers continue searching when their search cost is below this cutoff and stop when it is above it.

To simplify notation in the demand equations, we suppress the time subscript on the sampling probabilities ϕ.

A consumer purchases from the lowest-ranked store only if that store is drawn first and the consumer's search cost is high enough to stop immediately. Since the probability of sampling store $j = 1$ on the first draw is ϕ_{i1}, the demand share for the lowest-utility store in period t is

$$q_{i1}^t = \phi_{i1} \left[1 - G\left(c_{i1}^t\right)\right] \tag{6}$$

[2]This utility rank is not the order in which stores are sampled; it is the position of a store's realized indirect utility relative to the other store alternatives available to type i in period t.

[3]The number of available stores N_i may vary by consumer type because households shop in different markets. The utility ranking is indexed by t because prices, promotions, and availability change over time.

where G is the cumulative distribution function of search costs. Note that no type-i consumer whose search cost is below c_{i1}^t purchases from store 1 because the expected gain from another draw exceeds the search cost.

Following the same type of reasoning, as in Hortaçsu and Syverson (2004), we obtain the demand for the second lowest indirect utility store (numbered 2) as

$$q_{i2}^t = \phi_{i2}\left[1 + \frac{\phi_{i1} G\left(c_{i1}^t\right)}{1-\phi_{i1}} - \frac{G\left(c_{i2}^t\right)}{1-\phi_{i1}}\right] \tag{7}$$

and for stores ranked $j = 3, .., N_i$ in the indirect utility ranks:

$$q_{ij}^t = \phi_{ij}\left[\sum\nolimits_{k=1}^{j} \frac{G\left(c_{ik-1}^t\right) - G\left(c_{ik}^t\right)}{1-\sum_{l=0}^{k-1}\phi_{il}}\right] \tag{8}$$

which can be re-written as

$$q_{ij}^t = \phi_{ij}\left[1 + \sum\nolimits_{k=1}^{j-1} \frac{\phi_{ik} G\left(c_{ik}^t\right)}{\left(1-\sum_{l=0}^{k}\phi_{il}\right)\left(1-\sum_{l=0}^{k-1}\phi_{il}\right)} - \frac{G\left(c_{ij}^t\right)}{1-\sum_{l=0}^{j-1}\phi_{il}}\right] \tag{9}$$

where by convention $G\left(c_{i0}^t\right) = 1$ and $\phi_{i0} = 0$.

Equation (9) generalizes the same logic to stores of arbitrary rank j. Demand for a store reflects consumers who draw it directly and consumers who arrive there after drawing lower-ranked alternatives and continuing search. The final term subtracts consumers who reach rank j but still find it worthwhile to search again. This recursive structure is what allows the distribution of search costs to be recovered from purchase shares and sampling probabilities, without direct observation of the complete search path.

4.2 Discussion of Assumptions

When shopping at supermarkets, consumers typically purchase a basket of goods rather than a single item. We do not model the household's complete basket-search problem. Instead, the structural model separates two objects: an empirical store-exposure process for a household type in a region-week market, constructed from recorded purchase trips across food categories, and demand for products within the category under study, identified from purchases in that category.

This separation uses different sources of variation and relies on a simple interpretation of store visits. Store-exposure probabilities are constructed from recorded purchase trips across all food categories for similar households in the same region, week, and household type, excluding the household's own visits, as described in Section 4.3. A household observed buying milk at one store, pasta at a second, and coffee at a third therefore contributes information about the stores visited by households of its type, even though only purchases in the category under study enter that category's demand equation. The model does not require that a shopping trip be motivated by that category alone. It requires only that, conditional on visiting a store, the household can observe the prices and utility-relevant attributes of product variants in that category, so that the within-store choice in equation (1) and the store-level indirect utility in equation (2) are well defined regardless of what else the household purchased on the same trip.

This design lets us estimate category-specific demand and the price coefficient without specifying how a household allocates a full grocery basket across stores or how information acquired for one item affects purchases of another. This category-level interpretation is especially plausible for the products studied here. Coffee purchases may be triggered by household stockouts, while beer and whisky are often purchased for specific occasions. These products can therefore motivate store choice or price comparison on their own, even though many purchases still occur during broader grocery trips. Models of supermarket basket search, such as Wildenbeest (2011), and models of multi-category supermarket competition, such as Thomassen et al. (2017), are designed for questions where the basket itself is the central object. Our approach is complementary: we focus the search problem on a single product category, while using the recorded purchase diary across categories to measure the store-exposure process. The cost of this choice is that we do not model how baskets are split across stores or how search for one item affects search for another; addressing these questions would require specifying the full multiproduct search problem.[4]

Because the search decision involves visiting stores rather than only checking posted prices, the estimated search cost has a broad interpretation: it is the cost of acquiring and acting on

[4]See Carlson and McAfee (1984) and Zhou (2014).

price information. This includes transportation costs, time spent traveling, locating products within stores, processing price information, and any cognitive costs associated with evaluating alternatives. These activities are part of the cost of sampling an additional store; the model does not introduce a separate within-store search margin once the store has been sampled. We do not separately identify these components; instead, we estimate their combined effect as a single structural object.

We follow the empirical search literature in assuming perfect recall: once a consumer has sampled a store, the best previously sampled option remains available. With a finite number of stores, this assumption is useful because it keeps the optimal stopping rule stationary. Without perfect recall, the value of an additional search would depend on the full search history and on how many stores remain unsampled (Janssen and Parakhonyak, 2014), adding state variables that are not observed in our data. The assumption is also reasonable for the object we model. A consumer need not remember every observed price or indirect utility; she only needs to retain the best option encountered so far. Moreover, part of the search cost may be incurred when first locating the relevant product, registering its price, and comparing it with the current best alternative. Returning to a store already sampled may therefore involve travel and shopping time, but not necessarily the full information-acquisition cost paid on the first visit.

In our model, consumers search sequentially across stores. After each store draw, they decide whether to continue searching by comparing the cost of an additional draw with the expected benefit of finding a better alternative. An alternative is non-sequential, or fixed-sample, search. Under fixed-sample search, consumers decide before leaving home how many stores to visit before making a purchase decision, so the number of price draws is independent of the realizations encountered during search. De los Santos et al. (2012) test between sequential and fixed-sample search using data on web browsing and purchase behavior. They find that fixed-sample search better describes how consumers search for books online. For brick-and-mortar supermarkets, the sequential protocol is natural: a household that encounters a sufficiently attractive offer can buy during that visit and stop searching, rather than maintaining a fixed plan to inspect a predetermined number of stores.

The model also distinguishes persistent store heterogeneity from search frictions. We do not

interpret concentrated shopping, by itself, as evidence of costly search. Persistent differences across stores, such as location, convenience, assortment, loyalty, service quality, and other non-price attributes, enter the store-level indirect utility component. In the empirical specification, this component includes chain effects, time effects, household-type controls, and rural-by-chain interactions. Stores are therefore ranked according to the expected indirect utility of the store-category alternative, not according to price alone. Conditional on these rankings and on the estimated store-exposure probabilities, the search model asks whether observed purchases are consistent with consumers reaching higher-utility alternatives before purchase, or instead with consumers sometimes stopping after inspecting only a limited subset of stores.

4.3 Identification

To build intuition, suppose first that the econometrician observes the sampling probabilities ϕ_{ij} and the ranking of stores in indirect utility space. In this case, equations (6)–(9) map observed purchase shares into the values of the search cost distribution at the cutoff points, $G(c_{ij}^t)$, for $j = 1, \ldots, N_i$.

The main econometric challenge is that the utility ranking of stores is not directly observed. Under pure vertical differentiation, prices alone would determine rankings. In our setting, however, utilities depend on product prices, product characteristics, chain effects, region effects, and household type. Rankings must therefore be recovered from observed behavior. The structure of the model implies that the ratio

$$\frac{q_{is}^t}{\phi_{is}^t}$$

is strictly increasing in the latent utility of store s: stores that attract disproportionately many purchases relative to their probability of being sampled must occupy higher positions in the utility ranking.

For each market and household type, let $j(s)$ denote the rank of store s in utility space, ordered from lowest to highest utility. We recover this rank from

$$j(s) = 1 + \sum_{r=1}^{N_i} 1\left\{\frac{q_{ir}^t}{\phi_{ir}^t} < \frac{q_{is}^t}{\phi_{is}^t}\right\}. \tag{10}$$

This rank recovery is the analogue of the price ordering in Hortaçsu and Syverson (2004), but

it allows rankings to depend on all components of store utility rather than on price alone.

The empirical counterpart of the sampling probabilities, ϕ_{is}^t, is constructed from recorded purchase trips in all product categories, not only from purchases in the product category being modeled; a recorded trip is a visit on which the household bought at least one product. For each buyer in that category, we compute the probability of exposure to store s from the recorded visits of other households in the same region-week-household-type cell, excluding the household's own visits. The market-level probability is the average of these leave-one-household-out exposure probabilities across buyers in that category-market.

This leave-one-household-out construction removes the mechanical reflection problem that would arise if a household's own visits entered its own exposure measure. It does not, by itself, rule out common local factors that affect both store visits and store utility, such as access, parking, assortment, service quality, or persistent chain-region appeal. We therefore interpret ϕ_{is}^t as an empirical measure of store exposure rather than as a separately identified primitive search technology. The structural estimates should be read as conditional on this exposure process: given the stores that households of a given type are likely to encounter, the model uses category purchases to recover utility rankings, search behavior, and the implied substitution patterns.

Once stores are ranked by indirect utility, the store-exposure probability and purchase share associated with utility rank j are obtained by assigning each observed store to its recovered rank. Conditional on this utility ranking and store-exposure probabilities, the demand equations identify the values of the search cost distribution recursively. Let $\tilde{q}_{ij}^t = q_{ij}^t/\phi_{ij}^t$ and set $\tilde{q}_{i0}^t = 0$. Then

$$G(c_{ij}^t) = 1 - \sum_{k=1}^{j} \left(\tilde{q}_{ik}^t - \tilde{q}_{i,k-1}^t\right) \left(1 - \sum_{l=1}^{k-1} \phi_{il}^t\right). \tag{11}$$

The estimated values $G(c_{ij}^t)$ have a direct interpretation as the share of consumers who search beyond rank j: $1 - G(c_{i1}^t)$ is the mass of consumers who stop after one draw, while $G(c_{ij}^t) - G(c_{i,j+1}^t)$ is the mass who search beyond rank j but stop once they reach rank $j+1$ or better.

For structural estimation, we work with a finite number of utility ranks. The nonparametric recovery first keeps the original purchase shares and sampling probabilities for the modeled

ranks, so the remaining tail probability can be reported as a diagnostic of the truncation. The utility-recovery step then renormalizes purchase shares and store-exposure probabilities within the modeled rank set and recomputes G for that rank set, so the structural equations are internally coherent. The baseline specification uses five ranks, and we use four and six ranks as robustness checks.

Assuming that $G(\cdot, \theta)$ belongs to a known strictly monotone family, the cutoff levels are recovered through inversion:

$$c_{ij}^t(\theta) = G^{-1}\left(G(c_{ij}^t), \theta\right).$$

We use a lognormal distribution for search costs, with parameters $\theta = (\mu, \sigma)$. The recovered cutoff structure then allows us to back out indirect utilities, up to normalization, using equation (5).

The final step estimates the parameters of the within-store product utility in equation (2). Let d_{ijt} collect observed store, time, and household-type controls, including the rich and rural indicators and rural-by-chain interactions used in the baseline specification, and write the common component as $d_{ijt}\lambda$. The recovered utility satisfies

$$u_{ijt}(\theta) = d_{ijt}\lambda + \log\left[\sum_{b \in B_{jt}} \exp\left(x_{jbt}\beta - \alpha p_{jbt}\right)\right] + v_{ijt}. \tag{12}$$

Because product prices may be correlated with the unobserved utility component v_{ijt}, we estimate this equation using instrumental variables z_{ijt} satisfying

$$E[v_{ijt} z_{ijt}] = 0.$$

The baseline price instrument uses leave-one-out prices paid by other household types for the same narrowly defined product in the same retail chain and region-week, together with the exogenous controls. These product-level prices are averaged over the modeled products available in the store to form the store-level instrument used in the within-store indirect-utility equation. The instrument exploits common chain-region-week pricing and cost conditions. Prices paid by other demographic groups for the same product in the same chain-region-week are informative about the price faced by the focal household type because retailers set prices at the chain-region

level and because common wholesale costs, promotions, and local competitive conditions shift prices for all household types. At the same time, the leave-one-out construction excludes the focal household type's own purchases from the instrument.

The exclusion restriction is that, after conditioning on the included product, chain, calendar-month, year, rich, rural, and rural-by-chain controls, prices paid by other household types are not correlated with unobserved own-type demand shocks in v_{ijt}. In other words, the identifying variation comes from common pricing and cost shocks that move the product-level prices entering the focal type's within-store inclusive value, not from unobserved shocks to the focal type's preference for a particular chain-region alternative. This assumption allows other-type prices to shift the focal type's price-inclusive store utility while excluding them from the focal type's unobserved utility component.

Substituting equation (12) into the recovered utility expression yields the moment condition

$$E\left[\left(u_{ijt}(\theta) - d_{ijt}\lambda - \log\left[\sum_{b\in B_{jt}} \exp\left(x_{jbt}\beta - \alpha p_{jbt}\right)\right]\right) z_{ijt}\right] = 0. \tag{13}$$

Identification requires that the excluded instrument provide variation in product prices, and hence in the within-store inclusive values, conditional on the included controls. In particular, the leave-one-out price instrument must not be collinear with the chain, time, product, household-type, and rural-by-chain controls. Under this relevance condition and the exclusion restriction described above, the moment conditions identify the search cost parameters, product utility parameters, common store components, and the common price coefficient.

5 Estimation and Empirical Results

This section presents the structural estimates for the coffee category and the price elasticities implied by the model. We begin with the baseline specification, reporting the recovered search behavior, search-cost cutoffs, and preference estimates. We then use these estimates to compute own- and cross-store price elasticities, using the closed-form expressions derived in Appendix B. The discussion focuses on the economic interpretation of the results: how many stores consumers search before purchasing, how large the implied search costs are, and how these

costs shape substitution across supermarket chains.

5.1 Search Cost and Preferences Estimates

The structural estimation focuses on coffee because it provides the strongest support for estimating the full search-demand model. The model requires variation both across stores and within stores: enough store alternatives in a region-week-household-type market to recover search behavior, and enough purchases of narrowly defined products within the category to estimate the within-store product-choice equation. Coffee satisfies these requirements better than the other categories considered. Applying the same rank-5 requirement to each category, coffee has 4,368 region-week-household-type markets with at least five observed chain alternatives, compared with 3,106 for cola, 2,619 for beer, and 652 for whisky. After restricting the within-store product-choice equation to the twenty most frequently purchased coffee products, the final coffee estimation sample contains 3,585 rank-5 markets.

Throughout the structural analysis, coffee is the product category. The products entering the utility equation are narrowly defined coffee items constructed from scanner characteristics. The twenty most frequently purchased coffee products enter the within-store inclusive value; other coffee products are included in the category outside option.

The search object is a store in a region-week market. When a consumer samples a store, she observes the prices and characteristics of the coffee products included in the within-store product-choice equation and available there, and chooses her preferred product within the category. The store-level indirect utility that enters the search model is therefore the expected value of this within-store product choice. It is not an aggregate coffee price at the store. In the empirical implementation,

$$u_{ijt} = \delta_{ijt} + \log\left[\sum_{b \in B_{jt}} \exp\left(x_{jbt}\beta - \alpha p_{jbt}\right)\right] + v_{ijt}, \tag{14}$$

where B_{jt} is the set of modeled coffee products observed in store j in period t, p_{jbt} is the product-level price, and x_{jbt} denotes observed product characteristics. In the baseline specification, the common component δ_{ijt} includes chain indicators, calendar-month indicators, year indicators, rich and rural indicators, and rural-by-chain interactions. The price coefficient is common

across household types. This specification keeps the distinction between the product category, the narrowly defined products within the category, and the stores over which consumers search, while allowing store preferences and store-exposure probabilities to vary with household type.

Household heterogeneity enters through four observable types. We classify households by whether they live in a low-density area and whether household income is above the sample median, yielding urban poor, urban rich, rural poor, and rural rich households. Smaller chains are grouped into an Other chain category; this is a grouping of retail chains, not the outside option in the coffee product-choice equation.

For each region-week-household-type market, we construct coffee purchase shares across stores. Store-exposure probabilities are constructed from recorded purchase trips in all product categories, where a recorded visit is a trip on which the household bought at least one product. As described above, these probabilities are computed from other households in the same region-week-household-type cell, excluding the household's own visits. The baseline specification uses five utility ranks, with four and six ranks used as robustness checks. For the utility equation, price instruments are built from leave-one-out prices paid by other household types for the same product in the same chain and region-week. These product-level instruments are averaged across the narrowly defined coffee products available in the store and identify the price coefficient in the within-store indirect-utility equation.

Table 6: Share of consumers who search beyond each store rank

Consumer type	G_1	G_2	G_3	G_4	G_5
Urban poor	0.419	0.268	0.163	0.078	0.017
Urban rich	0.476	0.303	0.181	0.092	0.027
Rural poor	0.406	0.246	0.142	0.068	0.017
Rural rich	0.429	0.263	0.155	0.075	0.020

Notes: Entries are averages across market-week cells for each household type. G_j is the share of consumers who search beyond rank j, where stores are ranked from lowest to highest indirect utility. Thus, $1 - G_1$ is the share that stops after one store draw and $G_j - G_{j+1}$ is the share that stops once rank $j + 1$ or a higher-ranked store is reached. G_5 measures the residual share that searches beyond the fifth modeled rank.

Table 6 reports the share of consumers who search beyond each store rank. The shares decline rapidly with rank for all household types. For example, $1 - G_1$ is the share that stops after one store draw, and $G_1 - G_2$ is the share that searches beyond the first rank but stops

after reaching the second rank. Depending on household type, between 52 and 59 percent of consumers stop after the first store draw, while an additional 15 to 17 percent stop after reaching the second rank. Only 1.7 to 2.7 percent search beyond the fifth modeled rank. This small residual share supports the rank-5 baseline: the five modeled ranks account for nearly all search behavior in the coffee estimation sample. The main pattern is similar across household types: most coffee purchases occur after consumers observe only a small number of stores.

Table 7: Search-cost cutoffs by household type

Panel A. Utility units					
Consumer type	c_1	c_2	c_3	c_4	c_5
Urban poor	0.698	0.433	0.288	0.175	0.068
Urban rich	0.825	0.486	0.312	0.193	0.086
Rural poor	0.676	0.402	0.261	0.160	0.068
Rural rich	0.727	0.428	0.278	0.171	0.075
Panel B. Percent of average coffee price/kg					
Consumer type	c_1	c_2	c_3	c_4	c_5
Urban poor	15.0	9.3	6.2	3.8	1.5
Urban rich	17.7	10.4	6.7	4.1	1.8
Rural poor	14.5	8.6	5.6	3.4	1.5
Rural rich	15.6	9.2	6.0	3.7	1.6

Notes: $c_j = F^{-1}(G_j; \mu, \sigma)$ is the search-cost cutoff implied by the share of consumers who search beyond rank j and the estimated lognormal search-cost distribution. Panel B divides the euro/kg-equivalent cutoff by the average store-level coffee price in the estimation sample, 7.70 euros/kg. Search costs should be interpreted broadly as the cost of acquiring and acting on price information, including travel and shopping-trip frictions. Utility-unit cutoffs in Panel A are the main model-implied objects; Panel B provides an interpretable price-unit conversion using the estimated price coefficient.

The estimated lognormal distribution of search costs has parameters $\mu = -0.217$ and $\sigma = 1.050$. The distribution is dispersed: the median search cost is 0.805 in utility units, while the 90th percentile is 3.092. These values are expressed in units of store-level indirect utility and should be read as costs of acquiring and acting on additional price information. They combine travel time, store access, shopping-trip frictions, and cognitive price discovery; the model does not separately identify these components. The full distributional parameter table is reported in Appendix Table A1.

Table 7 translates the shares in Table 6 into search-cost cutoffs. For rank j, the cutoff is

$$c_j = F^{-1}(G_j; \mu, \sigma),$$

Table 8: Preference parameter estimates

Parameter	Estimate	Std. err.	t-stat.
Price coefficient α	0.606	0.082	7.38
Rich household	0.411	0.047	8.81
Rural household	0.118	0.070	1.69
Rural × chain 2	0.743	0.092	8.11
Rural × chain 3	0.401	0.098	4.11
Rural × chain 4	-0.801	0.149	-5.37
Rural × chain 5	-1.133	0.092	-12.34
Rural × chain 6	-1.188	0.099	-11.94

Notes: The table reports estimates from the within-store indirect-utility equation for coffee. The price coefficient is common across the four household types used to recover search behavior and store-exposure probabilities. Reported geography controls are interactions between the rural indicator and chain indicators; chain 1 is the omitted category. The specification also includes chain indicators, calendar-month indicators, year indicators, and product-level controls, which are not reported. The excluded price instrument is the leave-one-out average price paid by other household types for the same product and retail chain in the same region-week, averaged over modeled products available in the store. Standard errors are heteroskedasticity-robust sandwich standard errors conditional on the recovered search behavior and empirical store-exposure probabilities.

where F is the estimated lognormal search-cost distribution. Economically, c_j is the cost at which a consumer who has reached rank j is indifferent between stopping and drawing another store. The data identify the shares G_j, while the distributional assumption and the normalization of utility determine the cardinal cost scale. Panel B expresses each cutoff as a percentage of the average coffee price per kilogram, using the estimated price coefficient in the indirect-utility equation. These price-unit conversions are interpretive; the utility-unit cutoffs in Panel A are the main model-implied objects. For the first search cutoff, the implied values range from about 14.5 to 17.7 percent of the average coffee price per kilogram across household types. The euro/kg-equivalent cutoffs are reported separately in Appendix Table A4; a literal euro equivalent for a purchase occasion would additionally require multiplying by the relevant quantity of coffee purchased.

Table 8 reports the estimated price coefficient and the main non-price preference shifters. The estimate of α is 0.606, with a standard error of 0.081, so the marginal disutility of income is precisely estimated. The baseline specification uses one price coefficient for all household types. Household heterogeneity enters through the empirical store-exposure probabilities, the recovered search behavior, and the rich, rural, and rural-by-chain terms in store-level indirect

utility. This keeps the price coefficient well identified while preserving the geographic and demographic heterogeneity that matters for search across physical stores.

The estimated rich-household coefficient is positive and precisely estimated. The rural indicator is positive but only marginally significant. The rural-by-chain interactions are economically large and precisely estimated, showing that rural households value chains differently from urban households. This is the main source of observed household heterogeneity in store preferences in the baseline specification, alongside the type-specific store-exposure probabilities and search behavior.

We also examine whether the results depend on how store-exposure probabilities are constructed. The baseline measure uses all recorded purchase trips by other households of the same type in the same region-week. As robustness checks, we reconstruct exposure using only trips that do not include a coffee purchase, using lagged non-coffee trips from the previous eight weeks, and using a shrinkage rule that averages lagged non-coffee exposure with uniform exposure over available chain-region stores.[5] The non-coffee and lagged non-coffee specifications deliver price coefficients of 0.660 and 0.659, respectively, close to the baseline estimate of 0.606, and full own-store elasticities of -0.419 and -0.400, compared with -0.396 in the baseline. The shrinkage specification attenuates the price coefficient to 0.155, but it remains statistically significant. Appendix Table A2 reports the full set of results.

5.2 Price Elasticities

Having estimated the search cost distribution and consumer preferences, we compute the own- and cross-price elasticities of demand implied by the search model. Although search decisions are described in terms of a consumer's ordered set of store utilities, the relevant competitive objects are store-pair elasticities: the response of demand for store s to a price change at store l. This requires mapping the rank-based search problem back to stores, since the same store may occupy different positions in the utility ordering across markets and periods.

[5] A uniform exposure rule is not reported because it is not a credible measurement rule in this setting: households are unlikely to face equal exposure to all chain-region stores. Our robustness checks instead reduce reliance on contemporaneous coffee trips while preserving observed heterogeneity in store access.

In the search framework, prices affect demand through three related channels. First, a price change affects the indirect utility of a store conditional on that store being sampled. Second, it changes the search cutoffs, and therefore the share of consumers who find it worthwhile to continue searching after a given draw. Third, because stores are ranked by indirect utility, a price change can also affect the probability that a given store occupies each position in the utility ordering. The elasticity calculation therefore allows purchase probabilities, the shares who stop after each draw, and the probabilities that stores occupy each rank to respond to price changes.

A useful feature of this rank-based structure is that the store-level elasticity has a closed-form decomposition. Store-level demand for type i in market t can be written as

$$q_{is}^t(\mathbf{p}_t) = \sum_{j=1}^{N_i} P(j(s) = j \mid \mathbf{p}_t)\, q_{ij}^t(\mathbf{p}_t),$$

where $P(j(s) = j \mid \mathbf{p}_t)$ is the probability that store s occupies utility rank j, and $q_{ij}^t(\mathbf{p}_t)$ is demand for the store ranked j. Differentiating with respect to the price of store l gives

$$\frac{\partial q_{is}^t(\mathbf{p}_t)}{\partial p_{lt}} = \sum_{j=1}^{N_i} P(j(s) = j \mid \mathbf{p}_t) \frac{\partial q_{ij}^t(\mathbf{p}_t)}{\partial p_{lt}} + \sum_{j=1}^{N_i} q_{ij}^t(\mathbf{p}_t) \frac{\partial P(j(s) = j \mid \mathbf{p}_t)}{\partial p_{lt}} \tag{15}$$

The first term is the response holding the rank assignment of stores fixed. The second term captures the induced movement of stores across utility ranks. Appendix A gives the full derivative.

We report two related elasticity measures. The first holds fixed the probability that each store occupies each rank, evaluated at the recovered baseline values. This conditional elasticity isolates the response of demand through store utilities and search stopping, taking the ranking structure as given. The second measure also allows a price change to reshuffle stores in the utility ordering. This full elasticity is our main measure of substitution and market power, because it captures not only how consumers respond to price changes within a given ranking, but also how price changes affect which store occupies each position in the utility ordering.

Because the full elasticity includes the effect of prices on the ranking of stores, it need not satisfy the same sign restrictions as an elasticity computed conditional on a fixed ranking.

In particular, the full cross-store elasticity can be negative even when stores are substitutes conditional on a given ranking, and the full own-store elasticity can be positive in some cells even though conditional own-store elasticities are negative. Such cases do not reflect complementarity across stores or upward-sloping demand conditional on information. They reflect the fact that a price change can alter not only the attractiveness of a store, but also the probability that each store occupies each position in the utility ordering.

Table 9: Coffee own and cross price elasticities

Elasticity type	Mean cond.	Mean full	Share-weighted cond.	Share-weighted full
Own-store	-2.063	-0.324	-1.589	-0.394
Cross-store	0.451	0.078	0.397	0.104

Notes: Own-store elasticities are responses to a price increase at the same store. Cross-store elasticities are responses at other stores in the same market. Conditional elasticities hold fixed the probability that each store occupies each utility rank; full elasticities include the induced change in those probabilities. Share-weighted entries weight each store-pair elasticity by the baseline coffee purchase share of the demand store.

Table 10: Full store-level coffee elasticities, aggregated by chain

Demand / price chain	Other	Chain 1	Chain 2	Chain 3	Chain 4	Chain 5
Other	-0.380	0.086	0.148	0.034	0.160	0.169
Chain 1	0.082	-0.236	0.062	-0.009	0.038	0.034
Chain 2	0.174	0.049	-0.382	0.017	0.056	0.062
Chain 3	0.074	-0.013	0.029	-0.158	0.024	0.025
Chain 4	0.116	0.063	0.074	0.031	-0.354	0.101
Chain 5	0.132	0.043	0.082	0.031	0.098	-0.382

Notes: The underlying elasticity is computed at the store-pair level in market-type-week cells and then averaged by the chain identity of the demand store and the price-changing store. Full elasticities include the induced change in the probability that each store occupies each rank. Rows index the chain of the demand store; columns index the chain whose price changes. The Other row and column group smaller chains outside the displayed chain set; they are not the outside option.

Table 9 shows that search frictions substantially dampen substitution across stores. When the probability that each store occupies each utility rank is held fixed, demand is responsive to price changes: the mean own-store elasticity is -2.063, and the share-weighted own-store elasticity is -1.589. Once price changes are allowed to affect stores' positions in the indirect-utility ordering, the implied responses become much smaller, with corresponding own-store elasticities of -0.324 and -0.396. Cross-store substitution follows the same pattern. The share-weighted cross-store elasticity falls from 0.397 in the conditional measure to 0.104 in the full measure. These estimates imply that consumers respond to relative price changes across stores, but that the scope for substitution is limited by incomplete price information and by

the fact that price changes also reshape the probability that each store occupies each position in the utility ordering.

Appendix Table A3 reports the distribution behind these averages. Conditional own-store elasticities are negative throughout the reported distribution, with a median of -1.650. Full own-store elasticities are centered closer to zero, with a median of -0.371, and are more dispersed because they also include the effect of price changes on the probability that each store occupies each utility rank; the upper quartile of the full own-store distribution is positive. Cross-store elasticities are positive at the median in both objects, but the median falls from 0.343 in the conditional measure to 0.119 in the full measure. The lower tail of the full cross-store distribution is negative. These sign reversals arise when the utility-rank reallocation component dominates the direct utility and cutoff response in small store-pair cells; they do not imply that stores are economic complements or that own-store demand slopes upward conditional on information.

Table 10 shows the store-pair elasticities by chain identity. The diagonal entries are negative for all displayed chain groups, while most off-diagonal entries are positive. The largest off-diagonal entries remain modest, below 0.18 in absolute value, and the few negative off-diagonal entries are small. This chain-level aggregation therefore reinforces the main pattern: imperfect price information weakens substitution across retailers even when the within-ranking response to price changes is substantial.

The price elasticity results point to limited substitution across supermarket chains in response to coffee price changes. This is the central market-power implication of the search model: retailers face relatively inelastic local demand not only because stores are differentiated, but also because consumers observe only a subset of current store-level prices before purchasing.

5.3 Utility Loss from Imperfect Price Information

The elasticities in Table 9 show that search frictions substantially dampen substitution across stores. We next quantify a choice-utility cost of imperfect price information. For each market-week and household type, stores are ranked according to that type's recovered indirect utility. We compare the indirect utility of the highest-ranked store among the five store alternatives

included in the rank-5 specification to the purchase-probability-weighted expected indirect utility under the estimated search process. The difference measures the loss, in units of store-level indirect utility, from not purchasing at the best alternative among these five stores. This object is narrower than a full welfare measure: it does not include search, travel, or information-processing costs incurred while sampling stores, and it is computed within the five-store choice set of the rank-5 specification.

Table 11 reports the results. The main object is the utility-unit loss. In these units, the loss is similar across household types, ranging from 1.86 for rural-rich households to 2.31 for urban-rich households, with an overall mean of 2.12. The cross-type similarity mirrors the relatively similar search-cost cutoffs in utility units, although the levels are not directly comparable because the objects measure different margins.

To give the scale of these utility losses, one can translate them into a price-equivalent change using the estimated price coefficient. Since prices are measured in euros per kilogram, dividing the utility loss by $\alpha = 0.606$ gives the uniform coffee price increase, in euros/kg, that would generate the same loss in indirect utility. The overall mean loss of 2.12 utility units corresponds to about 3.49 euros/kg, or 45 percent of the average coffee price in the estimation sample.

This amount is not how much households literally spend because of imperfect information. The loss in Table 11 is a choice-utility loss: it compares expected indirect utility under the estimated search process with the indirect utility of the best store among the five modeled alternatives. The conversion is therefore a scale normalization, useful for interpreting magnitude, rather than a direct measure of money left on the table or travel costs paid.

The estimated choice-utility losses point to a consumer-side consequence of imperfect price information in addition to the competition channel measured by the elasticities. Limited search weakens substitution across retailers, and it also lowers expected indirect utility relative to the best store alternative available to the consumer. Policies or technologies that make price comparison easier, such as clearer unit pricing, digital price-comparison tools, or better disclosure of current prices, can therefore operate through two margins: stronger competitive pressure across retailers and a lower cost of finding higher-utility store alternatives.

Table 11: Choice-utility loss from imperfect price information, by household type

Consumer type	Cells	Mean	Median
Urban poor	700	2.20	1.79
Urban rich	907	2.31	1.76
Rural poor	927	2.16	1.66
Rural rich	1,051	1.86	1.57

Notes: For each household type and market-week cell, the loss is the recovered indirect utility of the highest-ranked store among the five store alternatives included in the rank-5 specification minus expected indirect utility under the estimated search process, using the model-implied full purchase probabilities (q^{full}) described in Section 5.2. Utility units are the main object of comparison. The measure excludes search, travel, and information-processing costs incurred while sampling stores, as well as losses relative to stores outside the five alternatives in the rank-5 specification. Results using rank-conditional purchase probabilities (q^{rank}) are very similar.

6 Conclusion

This paper estimates consumer search costs and demand in French supermarkets using household-level scanner data, and shows that imperfect price information is a quantitatively important feature of retail grocery markets. We document substantial price dispersion for narrowly defined branded products, and show that stores frequently change position in the price distribution from week to week. Because the identity of the lowest-price retailer is unstable over time, consumers cannot rely only on past experience to identify where to shop.

The descriptive evidence points to search frictions in household grocery shopping. Several household characteristics associated with time constraints and shopping costs are systematically related to prices paid and store visits. Households observed visiting more stores are also more likely to purchase at the lowest observed market price. These patterns hold after controlling for store and time fixed effects and a rich set of household characteristics.

We develop and estimate a structural demand model with sequential consumer search across differentiated stores, using narrowly defined products within the modeled category to identify preferences. The framework separates store exposure from category demand: store-exposure probabilities are constructed from observed visits across all food categories, while the demand system is estimated using purchases in the modeled category only. This allows the model to exploit the full richness of the shopping diary without treating broad baskets as the search object.

The structural estimates show that search frictions are quantitatively important. In the

coffee estimates, most consumers purchase after observing only one or two store alternatives, and very few consumers search deep into the utility-ranked set of stores. The recovered search-cost cutoffs are economically meaningful, especially for the first additional store draw. These frictions also have important implications for substitution. Allowing price changes to move stores across the indirect-utility ranking reduces the magnitude of the share-weighted own-store elasticity from -1.589 to -0.396 and the share-weighted cross-store elasticity from 0.397 to 0.104. The model also implies a consumer-side cost of imperfect price information: expected indirect utility under the estimated search process is below the indirect utility of the best store alternative available to the consumer.

Taken together, the findings suggest that imperfect price information is an important source of local market power in retail markets. When consumers observe only a subset of available prices before purchasing, substitution across store chains depends not only on preferences and prices, but also on which store prices are observed during search and on how much consumers continue searching. This has implications for the analysis of retail competition, merger simulation, and the welfare effects of policies aimed at improving price transparency, such as unit pricing requirements or digital price comparison tools.

References

Abaluck, J. and A. Adams-Prassl (2021). What do consumers consider before they choose? Identification from asymmetric demand responses. *Quarterly Journal of Economics 136*(3), 1611–1663.

Allen, J., R. Clark, and J.-F. Houde (2014). The effect of mergers in search markets: Evidence from the Canadian mortgage industry. *American Economic Review 104*(10), 3365–3396.

Carlson, J. A. and R. P. McAfee (1984). Joint search for several goods. *Journal of Economic Theory 32*(2), 337–345.

Chandra, A. and M. Tappata (2011). Consumer search and dynamic price dispersion: An application to gasoline markets. *RAND Journal of Economics 42*(4), 681–704.

Conlon, C. T. and J. H. Mortimer (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics 5*(4), 1–30.

Crawford, G. S., R. Griffith, and A. Iaria (2021). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics 222*(1), 4–43.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review 102*(6), 2955–2980.

DellaVigna, S. and M. Gentzkow (2019). Uniform pricing in US retail chains. *Quarterly Journal of Economics 134*(4), 2011–2084.

Hitsch, G. J., A. Hortaçsu, and X. Lin (2021). Prices and promotions in U.S. retail markets. *Quantitative Marketing and Economics 19*(3), 289–368.

Honka, E. (2014). Quantifying search and switching costs in the US auto insurance industry. *RAND Journal of Economics 45*(4), 847–884.

Honka, E., A. Hortaçsu, and M. R. Wildenbeest (2019). Empirical search and consideration sets. In J.-P. Dubé and P. E. Rossi (Eds.), *Handbook of the Economics of Marketing*, Volume 1, pp. 193–257. Amsterdam: North-Holland.

Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics 119*(2), 403–456.

Janssen, M. C. W. and A. Parakhonyak (2014). Consumer search markets with costly revisits. *Economic Theory 55*(2), 481–514.

Kaplan, G. and G. Menzio (2015). The morphology of price dispersion. *International Economic Review 56*(4), 1165–1206.

Kaplan, G., G. Menzio, L. Rudanko, and N. Trachter (2019). Relative price dispersion: Evidence and theory. *American Economic Journal: Microeconomics 11*(3), 68–124.

Koulayev, S. (2014). Search for differentiated products: Identification and estimation. *RAND Journal of Economics 45*(3), 553–575.

Lach, S. (2002). Existence and persistence of price dispersion: An empirical analysis. *Review of Economics and Statistics 84*(3), 433–444.

Martin, S. (2024). Market transparency and consumer search—Evidence from the German retail gasoline market. *RAND Journal of Economics 55*(4), 573–602.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2023). Consumer search and prices in the automobile market. *Review of Economic Studies 90*(3), 1394–1440.

Pennerstorfer, D., P. Schmidt-Dengler, N. Schutz, C. Weiss, and B. Yontcheva (2020). Information and price dispersion: Theory and evidence. *International Economic Review 61*(2), 871–899.

Seiler, S. (2013). The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics 11*(2), 155–203.

Small, K. A. and H. S. Rosen (1981). Applied welfare economics with discrete choice models. *Econometrica 49*(1), 105–130.

Thomassen, Ø., H. Smith, S. Seiler, and P. Schiraldi (2017). Multi-category competition and market power: A model of supermarket pricing. *American Economic Review 107*(8), 2308–2351.

Ursu, R., S. Seiler, and E. Honka (2025). The sequential search model: A framework for empirical research. *Quantitative Marketing and Economics 23*(1), 165–213.

Wildenbeest, M. R. (2011). An empirical model of search with vertically differentiated products. *RAND Journal of Economics 42*(4), 729–757.

Zhou, J. (2014). Multiproduct search and the joint search effect. *American Economic Review 104*(9), 2918–2939.

A Derivations of the Search Model

A.1 Demand System

We derive the demand system following Hortaçsu and Syverson (2004), adapted to heterogeneous sampling probabilities.

Let $F_{it}(u)$ denote the distribution of indirect utilities:

$$F_{it}(u) = \sum_{k=1}^{N_i} \phi_{ik} \mathbf{1}\{u_{ikt} \leq u\}. \tag{16}$$

Define cutoff search cost:

$$c_{ij}^t = \int_{u_{ijt}}^{\bar{u}_{it}} (u - u_{ijt}) dF_{it}(u). \tag{17}$$

Given discreteness:

$$c_{ij}^t = \sum_{k=j+1}^{N_i} \phi_{ik}(u_{ikt} - u_{ijt}). \tag{18}$$

A.2 Cutoff Ordering

Because utilities are strictly ordered:

$$u_{i1t} < u_{i2t} < \cdots < u_{iN_it}, \tag{19}$$

cutoffs satisfy:

$$c_{i1}^t > c_{i2}^t > \cdots > c_{iN_i}^t = 0. \tag{20}$$

A.3 Demand for the Lowest Ranked Store

A consumer purchases store 1 only if: - it is drawn first (probability ϕ_{i1}) - search cost exceeds cutoff c_{i1}^t

Hence:

$$q_{i1}^t = \phi_{i1}(1 - G(c_{i1}^t)). \tag{21}$$

A.4 Higher Ranked Stores

For store 2:

$$q_{i2}^t = \phi_{i2} \left[1 + \frac{\phi_{i1} G(c_{i1}^t)}{1 - \phi_{i1}} - \frac{G(c_{i2}^t)}{1 - \phi_{i1}} \right]. \tag{22}$$

For general j:

$$q_{ij}^t = \phi_{ij}\left[1 + \sum_{k=1}^{j-1} \frac{\phi_{ik} G(c_{ik}^t)}{(1-\sum_{l=0}^{k}\phi_{il})(1-\sum_{l=0}^{k-1}\phi_{il})} - \frac{G(c_{ij}^t)}{1-\sum_{l=0}^{j-1}\phi_{il}}\right]. \tag{23}$$

A.5 Recursive Identification of $G(\cdot)$

Define adjusted quantities:

$$\tilde{q}_{ij}^t = \frac{q_{ij}^t}{\phi_{ij}}. \tag{24}$$

Then:

$$\tilde{q}_{i1}^t = 1 - G(c_{i1}^t). \tag{25}$$

For $j = 2$:

$$G(c_{i2}^t) = 1 - q_{i1}^t - (1-\phi_{i1})\tilde{q}_{i2}^t. \tag{26}$$

In general:

$$G(c_{ij}^t) = 1 - \sum_{k=1}^{j}(\tilde{q}_{ik}^t - \tilde{q}_{ik-1}^t)\left(1 - \sum_{l=0}^{k-1}\phi_{il}\right), \tag{27}$$

with $\tilde{q}_{i0}^t = 0$.

A.6 Inversion to Search Costs

Assume $G(\cdot,\theta)$ is strictly monotone. Then:

$$c_{ij}^t = G^{-1}(G(c_{ij}^t),\theta). \tag{28}$$

A.7 Utility Recovery

Normalizing $u_{i1t} = 0$, utilities satisfy:

$$u_{i2t} = \frac{c_{i1}^t - c_{i2}^t}{\sum_{j=2}^{N_i}\phi_{ij}}. \tag{29}$$

For $k \geq 3$:

$$u_{ikt} = \frac{c_{i1}^t}{\sum_{j=2}^{N_i}\phi_{ij}} + \sum_{k'=2}^{k-1}\frac{\phi_{ik'}c_{ik'}^t}{(1-\sum_{l=0}^{k'}\phi_{il})(1-\sum_{l=0}^{k'-1}\phi_{il})} - \frac{c_{ik}^t}{\sum_{j=k}^{N_i}\phi_{ij}}. \tag{30}$$

A.8 Moment Conditions

Final estimation uses the recovered store utility and the within-store product logsum:

$$E\left[\left(u_{ijt}(\theta) - d_{ijt}\lambda - \log\left[\sum_{b\in B_{jt}} \exp\left(x_{jbt}\beta - \alpha p_{jbt}\right)\right]\right) z_{ijt}\right] = 0. \tag{31}$$

B Derivation of Price Elasticities in the Search Model

This appendix derives the store-level price elasticities used in Section 5. The search model is expressed in terms of utility ranks, while elasticities are reported at the store level. The derivation therefore maps rank-specific demand back to store identities and allows price changes to affect both search behavior and the probability that a store occupies each utility rank.

Let s and l denote store identities, and let j denote a utility rank. Ranks are ordered from lowest to highest utility, so rank 1 is the lowest-utility store and rank N_i is the highest-utility store available to consumer type i. Store-level demand for store s can be written as

$$q_{is}^t(\mathbf{p}_t) = \sum_{j=1}^{N_i} P_{isj}^t(\mathbf{p}_t) q_{ij}^t(\mathbf{p}_t), \tag{32}$$

where

$$P_{isj}^t(\mathbf{p}_t) \equiv P(j(s) = j \mid \mathbf{p}_t)$$

is the probability that store s occupies utility rank j, and $q_{ij}^t(\mathbf{p}_t)$ is demand for the store ranked j.

We consider a proportional price increase at store l. Let τ_l denote the corresponding log-price shifter, so that the price of each modeled product b at store l is multiplied by $\exp(\tau_l)$. The reported elasticity is therefore a derivative of log demand with respect to τ_l, evaluated at $\tau_l = 0$.

Differentiating equation (32) with respect to τ_l gives

$$\frac{\partial q_{is}^t(\mathbf{p}_t)}{\partial \tau_l} = \sum_{j=1}^{N_i} P_{isj}^t(\mathbf{p}_t) \frac{\partial q_{ij}^t(\mathbf{p}_t)}{\partial \tau_l} + \sum_{j=1}^{N_i} q_{ij}^t(\mathbf{p}_t) \frac{\partial P_{isj}^t(\mathbf{p}_t)}{\partial \tau_l}. \tag{33}$$

The first term is the response of rank-specific search demand holding rank probabilities fixed. The second term captures the induced movement of stores across utility ranks.

B.1 Price derivative of store utility

Let the deterministic component of store utility be

$$\bar{u}_{irt}(\mathbf{p}_t) = \delta_{irt} + \log\left[\sum_{b\in B_{rt}} \exp\left(x_{rbt}\beta - \alpha p_{rbt}\right)\right], \tag{34}$$

and let total store utility be

$$u_{irt} = \bar{u}_{irt}(\mathbf{p}_t) + v_{irt},$$

where v_{irt} is i.i.d. type I extreme value across stores.

Define

$$D^t_{ir,l} \equiv \frac{\partial \bar{u}_{irt}(\mathbf{p}_t)}{\partial \tau_l}.$$

Because τ_l shifts prices only at store l,

$$D^t_{ir,l} = -\mathbf{1}\{r = l\}\alpha \sum_{b\in B_{rt}} \omega_{irbt} p_{rbt}, \tag{35}$$

where

$$\omega_{irbt} = \frac{\exp\left(x_{rbt}\beta - \alpha p_{rbt}\right)}{\sum_{b'\in B_{rt}} \exp\left(x_{rb't}\beta - \alpha p_{rb't}\right)}$$

is the within-store product-choice probability implied by the logsum. Thus, a store-level proportional price increase changes the store's indirect utility by the price coefficient times the within-store expected product price.

B.2 Utility-rank probabilities

The utility-rank probabilities in equation (32) are computed from the distribution of store utility shocks. Let Π_i denote the set of all permutations of the N_i stores available to type i. A permutation

$$\pi = (\pi_1, \ldots, \pi_{N_i})$$

orders stores from highest to lowest utility, so π_1 is the highest-utility store and π_{N_i} is the lowest-utility store.

Under iid type I extreme value shocks, the probability of the complete ranking π is the

rank-ordered-logit probability

$$\mathcal{P}_i^t(\pi \mid \mathbf{p}_t) = \prod_{m=1}^{N_i-1} \frac{\exp\left(\bar{u}_{i\pi_m t}(\mathbf{p}_t)\right)}{\sum_{h=m}^{N_i} \exp\left(\bar{u}_{i\pi_h t}(\mathbf{p}_t)\right)}. \tag{36}$$

The last step of the ranking has probability one and is therefore omitted from the product.

Because the search model indexes ranks from lowest to highest utility, store s has rank j if and only if it appears in position $N_i - j + 1$ of the highest-to-lowest ordering π. Therefore,

$$P_{isj}^t(\mathbf{p}_t) = \sum_{\pi \in \Pi_i:\, \pi_{N_i-j+1}=s} \mathcal{P}_i^t(\pi \mid \mathbf{p}_t). \tag{37}$$

This expression is exact under the iid type I extreme value assumption. It does not require pairwise ranking events to be independent.

To differentiate the rank probability, define the conditional logit share of a remaining store at step m of ranking π:

$$\omega_{i\pi_h|m,t} = \frac{\exp\left(\bar{u}_{i\pi_h t}(\mathbf{p}_t)\right)}{\sum_{a=m}^{N_i} \exp\left(\bar{u}_{i\pi_a t}(\mathbf{p}_t)\right)}, \qquad h \geq m. \tag{38}$$

Differentiating the complete-ranking probability gives

$$\frac{\partial \mathcal{P}_i^t(\pi \mid \mathbf{p}_t)}{\partial \tau_l} = \mathcal{P}_i^t(\pi \mid \mathbf{p}_t) \sum_{m=1}^{N_i-1} \left[D_{i\pi_m,l}^t - \sum_{h=m}^{N_i} \omega_{i\pi_h|m,t} D_{i\pi_h,l}^t \right]. \tag{39}$$

The derivative of the probability that store s occupies rank j is then

$$\frac{\partial P_{isj}^t(\mathbf{p}_t)}{\partial \tau_l} = \sum_{\pi \in \Pi_i:\, \pi_{N_i-j+1}=s} \frac{\partial \mathcal{P}_i^t(\pi \mid \mathbf{p}_t)}{\partial \tau_l}. \tag{40}$$

B.3 Derivative of rank-specific search demand

We next differentiate demand for the store occupying rank j. Let

$$T_{ij} \equiv \sum_{r=j}^{N_i} \phi_{ir} = 1 - \sum_{r=0}^{j-1} \phi_{ir},$$

with the convention $\phi_{i0} = 0$. The demand equation for rank j can be written as

$$q_{ij}^t = \phi_{ij} \left[1 + \sum_{k=1}^{j-1} \frac{\phi_{ik} G(c_{ik}^t)}{T_{i,k+1} T_{ik}} - \frac{G(c_{ij}^t)}{T_{ij}} \right]. \tag{41}$$

The cutoff for rank j is

$$c_{ij}^t = \sum_{r=j+1}^{N_i} \phi_{ir} \left(\bar{u}_{irt}^{rank} - \bar{u}_{ijt}^{rank} \right), \tag{42}$$

where $\bar{u}_{ijt}^{rank}$ denotes the deterministic utility of the store occupying rank j at the baseline ranking.

Holding the rank assignment fixed, the derivative of the cutoff with respect to the price shifter at store l is

$$\frac{\partial c_{ij}^t}{\partial \tau_l} = \sum_{r=j+1}^{N_i} \phi_{ir} \left(D_{ir,l}^{rank,t} - D_{ij,l}^{rank,t} \right), \tag{43}$$

where $D_{ij,l}^{rank,t}$ is the derivative of the deterministic utility of the store occupying rank j with respect to τ_l. For a proportional price increase at a single store, this derivative is nonzero only for the rank occupied by the price-changing store.

Let $g(\cdot)$ denote the density of the search-cost distribution $G(\cdot)$. Differentiating equation (41) yields

$$\frac{\partial q_{ij}^t}{\partial \tau_l} = \phi_{ij} \left[\sum_{k=1}^{j-1} \frac{\phi_{ik} g(c_{ik}^t) \frac{\partial c_{ik}^t}{\partial \tau_l}}{T_{i,k+1} T_{ik}} - \frac{g(c_{ij}^t) \frac{\partial c_{ij}^t}{\partial \tau_l}}{T_{ij}} \right]. \tag{44}$$

B.4 Elasticities

Combining equations (33), (40), and (44) gives the full store-level derivative:

$$\frac{\partial q_{is}^t(\mathbf{p}_t)}{\partial \tau_l} = \sum_{j=1}^{N_i} P_{isj}^t(\mathbf{p}_t) \frac{\partial q_{ij}^t(\mathbf{p}_t)}{\partial \tau_l} + \sum_{j=1}^{N_i} q_{ij}^t(\mathbf{p}_t) \frac{\partial P_{isj}^t(\mathbf{p}_t)}{\partial \tau_l}. \tag{45}$$

Since τ_l is a log-price shifter, the full elasticity is

$$\varepsilon_{isl}^{full,t} = \frac{1}{q_{is}^t(\mathbf{p}_t)} \frac{\partial q_{is}^t(\mathbf{p}_t)}{\partial \tau_l}. \tag{46}$$

The conditional elasticity holds fixed the probability that each store occupies each utility rank and therefore uses only the first term of equation (45):

$$\varepsilon_{isl}^{cond,t} = \frac{1}{q_{is}^t(\mathbf{p}_t)} \sum_{j=1}^{N_i} P_{isj}^t(\mathbf{p}_t) \frac{\partial q_{ij}^t(\mathbf{p}_t)}{\partial \tau_l}. \tag{47}$$

The difference between $\varepsilon_{isl}^{full,t}$ and $\varepsilon_{isl}^{cond,t}$ is the utility-rank reallocation component. This component can be positive or negative because a price change can raise the probability that

a store occupies some ranks and lower the probability that it occupies others. Since demand differs across ranks when consumers stop searching, the utility-rank reallocation term is not sign-restricted. For this reason, full cross-store elasticities can be negative for some store pairs even when stores are substitutes conditional on a fixed utility ordering.

C Additional Empirical Tables

Table A1: Search-cost distribution, coffee

μ	σ	p10	p25	p50	p75	p90	N	Criterion	Iter.
-0.217	1.050	0.210	0.396	0.805	1.634	3.092	17925	0.0034	1453

Notes: Search costs are assumed lognormally distributed with parameters μ and σ. Percentiles report the implied distribution in utility units. The criterion is the minimized GMM objective for the current main specification.

Table A2: Robustness to alternative store-exposure measures

	Baseline	Non-coffee	Lagged non-coffee	Shrinkage
Markets	3,582	3,585	3,553	3,577
Share searching beyond rank 1	0.433	0.448	0.441	0.572
Share searching beyond rank 5	0.021	0.023	0.022	0.035
Price coefficient α	0.606	0.660	0.659	0.155
Std. error	(0.081)	(0.099)	(0.090)	(0.056)
Full own-store elasticity	-0.396	-0.419	-0.400	-0.342
Full cross-store elasticity	0.104	0.109	0.104	0.088
Mean choice-utility loss	2.120	2.538	2.292	0.945

Notes: The baseline exposure measure uses all recorded purchase trips by other households of the same type in the same region-week. The non-coffee measure excludes trips containing a coffee purchase. The lagged non-coffee measure uses non-coffee trips from the previous eight weeks. The shrinkage measure averages lagged non-coffee exposure with uniform exposure over available chain-region stores. Standard errors are reported in parentheses. Elasticities are share-weighted averages from the full search model.

Table A3: Distribution of store-level coffee price elasticities

Store pair	Elasticity object	N	Mean	p10	p25	p50	p75	p90
Own-store	Conditional	17925	-2.063	-4.010	-2.691	-1.650	-1.165	-0.891
Own-store	Full	17925	-0.324	-1.662	-1.013	-0.371	0.360	1.113
Cross-store	Conditional	71700	0.451	0.071	0.175	0.343	0.625	0.994
Cross-store	Full	71700	0.078	-0.489	-0.144	0.119	0.342	0.573

Notes: The unit of observation is a store-pair/market-type/week cell. Own-store rows vary the price of the same store. Cross-store rows vary the price of another store in the same market. Conditional elasticities hold fixed the probability that each store occupies each utility rank; full elasticities include the induced change in those probabilities.

Table A4: Implied search-cost cutoffs, euro/kg-equivalents

Consumer type	c_1/α	c_2/α	c_3/α	c_4/α	c_5/α
Urban poor	1.15	0.71	0.48	0.29	0.11
Urban rich	1.36	0.80	0.51	0.32	0.14
Rural poor	1.12	0.66	0.43	0.26	0.11
Rural rich	1.20	0.71	0.46	0.28	0.12

Notes: Entries divide utility-unit cutoffs by the estimated price coefficient α. Since product prices are measured in euros/kg, these are euro/kg-equivalents: the coffee price change that gives the same indirect-utility loss as the search cutoff. They are not literal out-of-pocket costs of visiting another store. The utility-unit cutoffs in Table 7 remain the main model-implied comparison across household types.