Markups Across the Income Distribution: Measurement and Implications

Kunal Sangani
Harvard*

December 5, 2022

Abstract

The Law of Diminishing Elasticity of Demand (Harrod 1936) conjectures that price elasticity declines with income. I provide empirical evidence in support of Harrod's conjecture using data on household transactions and wholesale costs. Over the observed set of purchases, high-income households pay 14pp higher retail markups than low-income households. Half of the differences in markups paid across households is due to differences in markups paid at the same store. Conversely, products with a high-income customer base charge higher markups: a 10pp higher share of customers with over $100K in income is associated with a 3–7 percent higher retail markup. A search model in which households' search intensity depends on their opportunity cost of time can replicate these facts. Through the lens of the model, changes in the income distribution since 1950 account for a 10–14pp rise in retail markups, with 30% of the increase since 1980 due to growing income dispersion. This rise in markups consists of within-firm markup increases as well as a reallocation of sales to high-markup firms, which occurs without any changes to the nature of firm production or competition.

*Email: ksangani@g.harvard.edu. First version: March 2022. I am grateful to Adrien Bilal, Gabriel Chodorow-Reich, Xavier Gabaix, Deivy Houeix, Erik Hurst, Kiffen Loomis, Pierfrancesco Mei, Namrata Narain, Pascual Restrepo, Peleg Samuels, Andrei Shleifer, Ludwig Straub, Lawrence Summers, Adi Sunderam, and participants at NBER SI 2022 Macro Perspectives and Micro Data / Macro Models for helpful comments. All errors are my own. This paper contains my own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the author and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

There is growing evidence that average markups in the U.S. economy have been rising (De Loecker et al. 2020, Barkai 2020, Autor et al. 2020, Gutiérrez 2017). A number of mechanisms put forward to explain this phenomenon relate the rise in markups to changes in the supply side of the economy, such as a decline in antitrust enforcement (Gutiérrez and Philippon 2018), the rise of superstar firms (Autor et al. 2017), and structural technological change (De Loecker et al. 2021).

How changes in the demand side of the economy may contribute to the rise in markups is less studied. Under Harrod's (1936) conjecture that price sensitivity declines in income, the price elasticities facing firms—and hence the optimal markups charged by firms—depend on the level and distribution of income. In particular, a shift in the composition of demand toward high-income households leads to a decline in aggregate price sensitivity and hence a rise in markups. In this paper, I provide empirical evidence that the behavior of retail markups is consistent with Harrod's conjecture and explore its implications for the evolution of markups over time.

This paper starts by establishing that higher-income households systematically pay higher retail markups. I construct a measure of retail markups by pairing Nielsen Homescan data on purchases by a panel of households with wholesale cost data from PromoData Price-Trak, which monitors weekly changes to prices charged by wholesalers to retailers across the U.S. The merged dataset covers 26 million transactions made in a single year, accounting for 43 percent of transactions and 37 percent of expenditures in the Nielsen Homescan data. Relative to using data from a single retailer, this merged dataset has two advantages: (1) since it includes all household expenditures in Nielsen-tracked categories, it captures patterns of substitution across retailers; and (2) it includes detailed demographic information for households, typically not available in standalone retailer data. As argued by Gopinath et al. (2011), since rent, capital, and labor are fixed at short horizons, it is natural to interpret the wholesale cost as the marginal cost facing the retailer. Hence, I calculate retail markups—hereafter referred to as markups for brevity—as the ratio of the transaction price to the product's wholesale cost in the month of purchase.

The data indicate stark differences in markups paid across income groups. Over the observed set of purchases, the average markup paid by households increases from 29 percent for households with $10,000 in annual income to 43 percent for households with over $200,000 in annual income (i.e., a 14pp markup gap). The gap in markups paid between high- and low-income households is robust to the inclusion of demographic controls including household size, age, and race/ethnicity.

To relax the assumption that local inputs, transport costs, or store-specific factors do not affect marginal cost, I explore differences in markups paid within county and within store. Three-quarters of the markup gap (12pp) persists within county, and half of the markup gap (7pp) persists within store. Measures of the elasticity of markups paid with respect to household income yield a similar conclusion: the elasticity of markups paid to household income within store is 2.0%, over half of the unconditional elasticity of markups to household income (3.1%).

These findings add to a literature that has documented systematic differences in prices paid for identical products across households with different incomes (e.g., Aguiar and Hurst 2007a, Broda et al. 2009, Handbury 2021). I show that the gap in markups paid is two times larger than the gap in prices paid for identical products. Differences in prices paid for identical products underestimate true differences in price sensitivity across households because high-income households opt to buy other, high-markup substitutes. The use of wholesale cost data to calculate markups uniquely allows for these cross-good comparisons.

In principle, differences in markups paid across income groups could arise without a systematic relationship between products' average markups and buyer income. I show, however, that products purchased by high-income households have higher retail markups on average. In magnitudes, a 10pp increase in the share of purchases of a product coming from households with over \$100,000 in income is associated with a 3–7 percent increase in the product's retail markup.[1] Moreover, the link between product markups and buyer income is not explained by supply-side factors such as differences in products' market shares, differences in product market concentration, differences in the extent of variety in product categories, and whether the product category is a necessity or luxury. In fact, the link between product markups and buyer income persists even within products sold by the same firm in the same product category.

Finally, I show that the markups paid by a household depend not only on the household's own income, but also on the income of other households—for example, households in the same metro area or other households buying the same product. Evidence from both the cross-section and from the time series suggests the elasticity of markups paid to others' incomes is positive and large, implying that the "macro" elasticity of markups to income—the percent change in the average markup that would result if all households' incomes double—is greater than the micro elasticity documented in the cross-section of households. While the cross-sectional elasticity is around 3 percent, I estimate that the macro elasticity of markups to income lies between 8–15 percent.[2]

Taken together, these observations provide empirical support for Harrod's (1936) Law of Diminishing Elasticity of Demand: less price-sensitive, high-income households pay higher retail markups, and firms thus optimally set higher markups on products purchased by high-income households. The implication is that markups charged by firms depend on the composition of demand and hence on the income distribution.

To investigate this mechanism formally, I develop a model of consumer search building on the nonsequential search model of Burdett and Judd (1983).[3] The model features households that differ in labor and search productivity, leading to different opportunity costs of time across households.[4]

[1]These results are consistent with interviews of pricing decision makers by Bewley (2007). From interviews of managers in supermarkets and department stores, Bewley (2007) summarizes: "Mark-ups tend to be larger on expensive items bought by wealthier customers, because it is assumed that these buyers are insensitive to price."

[2]Previous estimates from the trade and spatial literatures are similar. For example, using data from an online retailer, Simonovska (2015) finds that the elasticity of markups to per-capita income in the destination country is 12–24 percent. Recent work by Bhardwaj et al. (2022) uses the introduction of million dollar plants as a quasi-exogenous shock to wages and finds an elasticity of around retail prices to hourly wages of 10 percent.

[3]Online Appendix E shows that a sequential search model à la Burdett and Mortensen (1998) generates similar results. Online Appendix F presents a model in which differences in markups paid across households instead arise from differences in elasticities of substitution and taste for quality, as in Handbury (2021) and Faber and Fally (2022).

[4]Pytka (2018) and Nord (2022) also develop search models with heterogeneous households. I clarify the differences between this paper and their results in the literature review below.

The household decision on shopping effort parallels the canonical model of price search by Aguiar and Hurst (2007a), but the returns to search are determined in equilibrium by the search behavior of all other households and by firm profit maximization. As a result, households' search decisions and the distribution of markups charged by firms are endogenous, equilibrium outcomes.

Cross-sectional analysis of household search behavior across U.S. counties provides suggestive evidence for the model mechanism. The model predicts that search intensity is decreasing with income but that households optimally search more in high-income areas. The latter prediction arises because household search decisions are strategic substitutes in equilibrium. In keeping with this prediction, search intensity—measured as the number of shopping trips or unique stores visited per dollar spent[5]—is decreasing in income (as previously documented by Pytka 2018), but increasing in the average county income.

The model allows me to derive comparative statics of the aggregate markup with respect to the distribution of buyers' incomes. These comparative statics depend crucially on how the opportunity cost of time spent shopping varies with household income. I show that a first-order stochastic shift in the distribution of buyers' incomes increases the aggregate markup if this opportunity cost is increasing in income. (In principle, the opportunity cost of time spent shopping may instead decrease with income if low-income households have limited access to search technologies, thus generating a "poverty premium" for low-income households.) A mean-preserving spread increases the aggregate markup if the opportunity cost of time spent shopping is increasing and convex in income.

To understand whether these conditions hold in the data, I calibrate the model to match differences in the retail markups paid across income groups. The estimated search intensities decrease with income: high-income households retrieve on average 40 percent fewer price quotes per purchase than low-income households. Search productivity increases less than one-for-one with labor productivity and at a decreasing rate in income, satisfying both conditions above.

The calibration suggests significant spillovers of households' search behaviors on outcomes of other households. Low-income (high-income) households retrieve 10 percent more (2 percent fewer) price quotes on average than they would in an economy of homogeneous households of their income level. In terms of markups paid, low-income households pay 5–9pp higher markups than they would in an economy populated only by low-income households. In contrast, the highest income households in the sample save over 10pp in average markups paid due to the presence of low-income shoppers. In other words, the search effort exerted by low-income households generates a significant positive externality for high-income households.[6] Accordingly, the model is able to generate a macro elasticity of markups to income nearly three times larger than the micro

[5]As pointed out by Pytka (2018), it is important to normalize total time spent shopping by the size of the consumption basket to compare search intensity across income groups. In Online Appendix B, I show that results are similar if, instead of normalizing by dollar spent, I normalize by number of transactions, number of unique UPCs purchased, or number of unique brands purchased.

[6]These externalities are distinct from the shopping externalities modeled by Kaplan and Menzio (2016). There, shopping externalities operate through employment, since unemployed shoppers search more and therefore reduce revenues for all firms.

elasticity in the cross-section, consistent with the data.

I use the model to return to the question that opens this paper: How do changes in the income distribution affect the aggregate markup? I consider the U.S. distribution of post-tax real income from 1950–2018 documented by Saez and Zucman (2019). Holding all other factors constant, these changes in the income distribution account for a 10–14pp rise in the aggregate markup from 1950 to 2018 through the lens of the model. Increases in the aggregate markup are moderate before 1980, but accelerate from 1980–2018 due both to the rising level and rising dispersion of incomes. Over 30% of the rise in the aggregate markup predicted after 1980 is attributed to changes in income dispersion. Data on gross margins from the Census of Annual Retail Trade Statistics suggests the increase in markups predicted by the model is consistent with the data.

The rise in the aggregate markup predicted by the model occurs due to both a rise in markups charged by individual firms as well as a reallocation of sales to high-markup firms. High-markup firms expand because high-income households receive a small number of price quotes and therefore shop at the top end of the markup distribution. Quantitatively, the model predicts that firms at the lowest end of the markup distribution lose 15 percent in sales share from 1950 to 2018, while the sales shares of firms at the highest end of the markup distribution grow over 30 percent. In all, the calibration suggests that changes to the composition of demand can be a potent force in both reshaping market structure and in increasing the aggregate markup over time.

Related literature. This paper contributes to a rich literature documenting differences in prices paid by income. Aguiar and Hurst (2007a), Broda et al. (2009), Handbury (2021), and Pisano et al. (2022) show that prices paid for identical products systematically rise with household income.[7] The empirical evidence in this paper augments the literature by showing that the elasticity of markups paid to household income is two times greater than the elasticity of prices paid for identical products to income. I build upon the insights of Aguiar and Hurst (2007a) and Kaplan and Menzio (2015) that search and opportunity cost of time play an important role in prices paid.[8]

A related trade literature explores why export prices are correlated with per-capita income in the export destination. Using data from an online retailer, Simonovska (2015) shows that the price of identical goods across countries correlates with per-capita income, and that this pattern is not explained by shipping costs or local inputs. Alessandria and Kaboski (2011) argue that households' comparative advantage in search in low-income countries can explain differences in prices across export destinations. In this paper, the link between income, search, and markups generates differences in markups paid by households within an economy, rather than across countries as in Alessandria and Kaboski (2011). The robust empirical relationship between markups and per-capita income suggest that markups need not follow a balanced growth path.[9]

[7]These studies define identical products at various levels of disaggregation, the narrowest of which is by product barcode and the broadest of which is typically a Nielsen "product module."

[8]See also subsequent empirical work on search effort and the use of savings technologies including Griffith et al. (2009), Aguiar et al. (2013), Coibion et al. (2015), Pytka (2018), and Nevo and Wong (2019).

[9]Relatedly, Menzio (2021) explores how a balanced growth path may be compatible with declining search frictions due to increasing (endogenous) specialization by sellers. This paper makes the complementary point that rising search

The model of search and markups developed in this paper is most closely related to work by Pytka (2018) and Nord (2022), who also develop models in which households have heterogeneous search intensities. Both embed these households in an Aiyagari (1994) model to explore differences between expenditure and consumption inequality.[10] The analytic comparative statics with respect to the income distribution, the use of retail markup data to calibrate the model, and the application to changes in markups over time are novel to this paper.

A key empirical contribution of this paper is to document that the behavior of retail markups is consistent with Harrod's conjecture. Several papers offer complementary evidence.[11] Lach (2007) finds that the arrival of price-sensitive immigrants from the Soviet Union led to a decrease in prices in Israeli neighborhoods where they moved and provides evidence that these immigrants spent more time shopping. Stroebel and Vavra (2019) show that retail prices and markups increase when homeowners' housing wealth rises. Anderson et al. (2018) find that markups charged by a single retail chain covary with local income due to product assortment but not product prices. By regressing changes in product sales on instrumented price changes, DellaVigna and Gentzkow (2019) find that demand elasticities are lower for stores in higher-income areas, and Faber and Fally (2022) find that higher quantiles of household income have small but statistically significant differences in price elasticities. Handbury (2021) finds that non-homotheticities in the elasticity of demand and valuation of quality are important to match observed market shares. Most recently, Auer et al. (2022) identify differences in the elasticities of substitution of high- and low-income Swiss households using exchange rate shocks. Where possible, I discuss how my results on price sensitivity over the income distribution compare to these previous findings.

A recent industrial organization literature explores trends in markups over time by estimating demand systems in Nielsen scanner data. Rather than recover markups from demand estimation, I use price and cost data to estimate retail markups. Nevertheless, the mechanisms I explore resonate with this literature. Döpper et al. (2021) attribute rising markups in Nielsen scanner data to declining consumer price sensitivity, echoing the mechanism developed in this paper, and incomplete pass-through of marginal cost reductions. Brand (2021) also finds that consumers have become less price sensitive, perhaps due to increased product differentiation.

Finally, the analysis in this paper relates to studies that link the income distribution to markups using non-homothetic preferences, such as Fajgelbaum et al. (2011) and Bertoletti and Etro (2017).

Structure of the paper. Section 2 describes the data sources and how I construct a measure of retail markups. Section 3 presents the main empirical findings: high-income households pay higher markups, products sold to a high-income customer base tend to have higher markups, and markups paid depend positively on others' incomes both in the cross-section and time series.

productivity need not result in lower markups. Instead, it is the race between two productivities—the relative growth rates of labor and search productivity—that matters for the evolution of markups in the model.

[10] Nord (2022) also provides evidence that the skewness of price distributions in Nielsen data covaries with customers' incomes, consistent with predictions of a search model, and considers implications for markups over the business cycle.

[11] A related marketing literature also explores how income and other characteristics predict a household's likelihood to be "coupon-prone" (e.g., Dickson and Sawyer 1990, Hoch et al. 1995, and references therein).

Section 4 presents a model in which differences in markups paid arise from differences in search intensity. Section 5 provides suggestive evidence that search behavior in the data conforms with the predictions of the model. Section 6 calibrates the model and quantifies spillovers across households. Section 7 considers how changes in the income distribution over time affect the aggregate retail markup through the lens of the model. Section 8 concludes.

2 Data Construction

In this section, I describe two data sources—Nielsen Homescan and PromoData Price-Trak—that I use to construct measures of retail markups paid by households. Appendix A provides a detailed description of how the two datasets are cleaned and merged.

2.1 Data sources

Consumer panel data. Nielsen Homescan provides transaction data for a nationally representative group of households over the period 2004–2019. In the main text, I present results using data from 2007, which covers 62 million transactions by over 60,000 households in about 2,700 counties.[12] The product categories tracked by Nielsen cover about 30 percent of all expenditure on goods in the consumer price index.[13]

Nielsen Homescan panelists use in-home scanners or a mobile application to record all purchases intended for personal, in-home use. The data include all Nielsen-tracked categories of food and non-food items purchased at any retail outlet. In addition to reporting the date and store location of each shopping trip, panelists scan the universal product code (UPC) of each item purchased, report the number of units purchased, and record savings from coupons. While Nielsen does not pay panelists, it offers households a variety of incentives to accurately report data.

Nielsen also collects annual demographic data on panelists including the age of each household member, race, employment status, household size, and household income. Panelists report household income for the full calendar year prior to the start of the panel year (i.e., households in the 2007 panel report income from the full 2005 calendar year). As in Handbury (2021), I exclude households with below \$10K income or missing income data from my analysis.

I make use of Nielsen's product hierarchy, which organizes products into product groups and modules. There are about 125 product groups and just over 1,000 highly disaggregated product modules. For example, "jams, jellies and spreads" is a product group, which consists of nine product modules for jams, jelly, marmalade, preserves, honey, fruit spreads, peanut butter, fruit and honey butters, and garlic spreads.

[12] From 2006-2009, Nielsen Homescan separately identifies households with \$100K, \$125K, \$150K, and \$200K+ in household income. These distinctions are not available prior to 2006 or after 2009. As we will see, markups paid vary significantly with income in this range. I report results for 2007 to avoid the impact of the recession in 2008–2009. Results from other years are similar and are available upon request.

[13] See Broda and Weinstein (2010) for a detailed description of the Nielsen Homescan data.

Wholesale costs. I use data on wholesale costs from PromoData Price-Trak, a weekly monitoring service that tracks wholesale prices for over 100,000 UPCs. The PromoData come from 12 grocery wholesaler organizations that sell products to retailers across the U.S. and covers the period 2006–2012.[14] On a weekly basis, wholesalers send PromoData order prices and promotional discounts that they make available to their customers.[15] Previous studies using these data include Nakamura and Zerom (2010), Stroebel and Vavra (2019), and Afrouzi et al. (2021).

The wholesale cost data include both base prices and "deal prices." Deal prices are discounts offered to retailers during promotions. These deal prices are only available to retailers during windows scheduled by the wholesaler and may require retailers to provide proof of promotion in order to redeem the discounted price. I present results using base prices as the measure of retailers' wholesale costs. As I show in Online Appendix B, similar results obtain from instead using deal prices as the measure of wholesale costs.

Consistent with Stroebel and Vavra (2019), I show in Appendix Table A1 that wholesale costs are similar across markets: in each month, 80 percent of items have a wholesale cost exactly equal to the modal wholesale cost observed across markets in that month. Hence, I calculate a national wholesale price for each UPC in each month.[16] In all, about 67,000 UPCs purchased by Homescan panelists in 2007 are matched to wholesale costs from PromoData. These UPCs constitute 43 percent of transactions and 37 percent of expenditures in the 2007 panel. Appendix Table A3 shows that this match rate is similar across income groups and that the relative prices of unmatched products covary with income in the same direction as matched products.

Constructing retail markup estimates. I calculate the retail markup on product g purchased by household i in transaction t as the price paid by i over the wholesale cost of product g in the month of transaction t,

$$\text{Retail Markup}_{i,g,t} = \frac{\text{Price}_{i,g,t}}{\text{Wholesale cost}_{g,t}}.$$

This approach is similar to Gopinath et al. (2011) and Anderson et al. (2018), who measure retail markups as price over replacement costs for a single retailer.[17]

Wholesale costs serve as a reasonable proxy for marginal cost for two reasons. First, according to the Census's Annual Retail Trade Statistics, wholesale costs account for three-fourths of total retail costs (including operating expenses and overhead) and constitute the largest portion of cost of goods sold. Annual reports from public grocery companies record a similar proportion of total

[14] A significant portion of grocery retail sales pass through wholesalers. In 2007, grocery store retail sales totalled \$491B (Census of Retail Trade). Total sales by merchant wholesalers in grocery and related products in that year were \$476B (Census Monthly Wholesale Trade Survey), and about half of grocery wholesalers' sales (excluding those to other wholesalers) are to retailers (2012 Economic Census).

[15] PromoData prices include wholesaler markups, which PromoData estimates are 2–5% of list price.

[16] An alternative is to only use the subset of transactions where a wholesale cost is reported for the market where the purchase is made. Appendix Figure B5 and Table B2 replicate the main analyses of the paper using only this subset of transactions. The results are quantitatively similar to the baseline results.

[17] The replacement costs used by Gopinath et al. (2011) include wholesale costs and total allowances (which include freight and transportation costs and subtract net rebates) from a single retailer. I observe only listed wholesale costs, and not transportation or rebate allowances. Industry reports suggest freight costs constitute 2-5% of costs of good sold.

costs coming from wholesale costs.[18] Second, as argued by Gopinath et al. (2011), since rent, capital, and labor are fixed at short horizons, it is natural to understand the replacement cost as the full marginal cost faced by the retailer.

Nevertheless, the wholesale cost data may mismeasure the marginal costs faced by retailers for three reasons: (1) true wholesale costs may vary from the list prices recorded by PromoData due to retailer-wholesaler deals, such as negotiated rebates or volume discounts, (2) true replacement costs may differ from wholesale costs due to other components of replacement costs, such as freight and transportation costs, and (3) true marginal costs may differ from replacement costs, for instance due to local inputs such as labor for shelving and inventory management. The empirical analysis in Section 3 seeks to control for each of these potential sources of mismeasurement—for example, by comparing markups paid by households within a specific store location—to ensure that the results are robust to each.

Retail markups are winsorized at the 1 percent level for all analyses. The sales-weighted average markup in the merged Homescan-PromoData dataset is 32 percent. This is moderately higher than the average markup of 28 percent in data from Dominick's Finer Foods and moderately lower than 41 percent for food and beverage stores from the 2007 Census Annual Retail Trade Statistics.[19]

3 Empirical Evidence

This section presents three sets of analyses. First, I show that high-income households pay higher retail markups on average. Differences in markups paid within store—which plausibly control for idiosyncratic differences in marginal costs due to labor or rent—account for over half of the gap.

Second, I show that products consumed by high-income households tend to have higher markups. In the cross-section, a 10pp increase in the share of consumers with over $100K in household income is associated with a 3–7 percent higher markup.

Finally, I document that, conditional on income, markups paid by a household are also increasing in the income of other households. These spillovers generate a macro elasticity of markups to income that is higher than the micro elasticity observed in the cross-section. Evidence from both the cross-section and time series places the macro elasticity of markups to income around 8–15%.

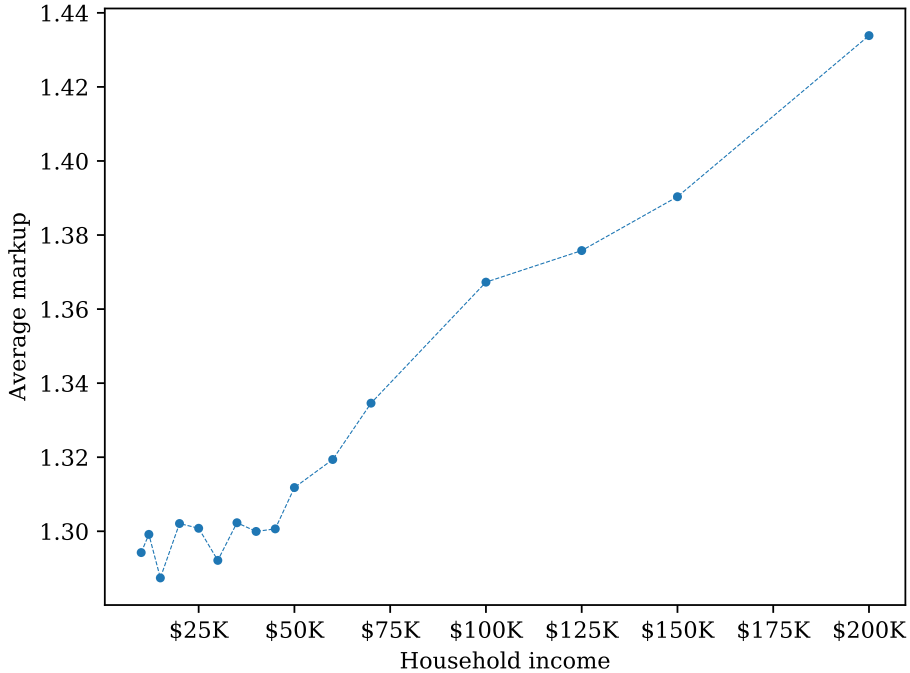
3.1 Fact 1. High-income households pay higher markups

Figure 1 plots the (unconditional) sales-weighted average markup paid by households over the income distribution for the set of products matched to wholesale costs. The average sales-weighted

[18]For example, in 2019, Kroger Co. reported merchandise costs—which include "product costs net of discounts and allowances; advertising costs; inbound freight charges; warehousing costs including receiving and inspection costs; transportation costs; and food production costs"—of $95.2M. Operating, general, and administrative expenses, which consist primarily of employee-related costs, and rent expense were $22.1M.

[19]The Census Annual Retail Trade Statistics reports a gross margin of 28.9 percent for food and beverage stores in 2007. Under constant returns to scale, this implies an aggregate markup of $1/(1 - 0.289) = 1.41$.

Figure 1: Sales-weighted average markup paid by income group.



markup paid by households increases from 29 percent for households with \$10–12K in annual income to 43 percent for households with over \$200K in annual income.[20]

To address concerns about mismeasurement of marginal cost mentioned above, I document the difference in markups paid between low- and high-income households within county and within store. Adding successive sets of controls absorbs factors that may lead to systematic differences in marginal cost across counties or across stores, such as differences in transportation costs or local input costs, thus isolating the differences in markups.

The first specification adds demographic controls and county fixed effects:

$$\text{Markup}_{i,g,t} = \sum_{\ell} \beta_{\ell} 1\{i \text{ has income level } \ell\} + \gamma' X_i + \delta_{\text{County}} + \epsilon_{i,g,t}. \tag{1}$$

The markup paid by household i for good g in transaction t is $\text{Markup}_{i,g,t}$. Demographic controls X_i include fixed effects for race, ethnicity, household size, presence of a female head of household, and the age group of the female head of household; δ_{County} are county fixed effects; and $\epsilon_{i,g,t}$ is a mean-zero error. I weight the regression by sales and leave out the income level indicator for households with less than \$20K income, so that the coefficients β_{ℓ} are differences relative to the below \$20K income group.

[20]The gap in the cost-weighted average markup paid across income groups is nearly identical at 14pp. I present the sales-weighted average here—and use sales-weighted regressions to decompose the markup gap—since sales are directly observed in the data, while costs are imputed. However, Appendix Table B1 and Figure B7 replicate the analyses from the main text using cost weights and find similar results.

Figure 2a plots the coefficients β_ℓ for specification (1) with and without county fixed effects. After accounting for demographic controls, the fixed effect on income for the highest-income group is 16pp, which is larger than the unconditional difference of 14pp. After adding county fixed effects, the fixed effect on income for the highest-income group is 12pp. Hence, over 70 percent of the difference in retail markups paid between the highest and lowest income groups in the sample is due to differences in markups paid within county.[21]

For just over half of the transactions in the sample, Nielsen provides store IDs that identify the specific store outlet where each purchase was made. Specification (2) adds store fixed effects for this subsample:

$$\text{Markup}_{i,g,t} = \sum_{\ell} \tilde{\beta}_\ell 1\{i \text{ has income level } \ell\} + \tilde{\gamma}' X_i + \alpha_{\text{Store}} + \epsilon_{i,g,t}. \tag{2}$$

Figure 2b plots the coefficients on income level for specification (2) with and without store fixed effects for the subsample of 14.0 million transactions where a unique store ID is available. For the subsample of transactions made at an identified store outlet, the fixed effect on income for the highest-income group is moderately smaller than in the full sample at 12pp.[22] Adding store fixed effects brings the difference in markups paid by the lowest and highest-income households to 7pp. Accordingly, for the sample of transactions made at a Nielsen-identified store, over half of the gap in markups paid by low- and high-income households is due to differences in markups paid within store.

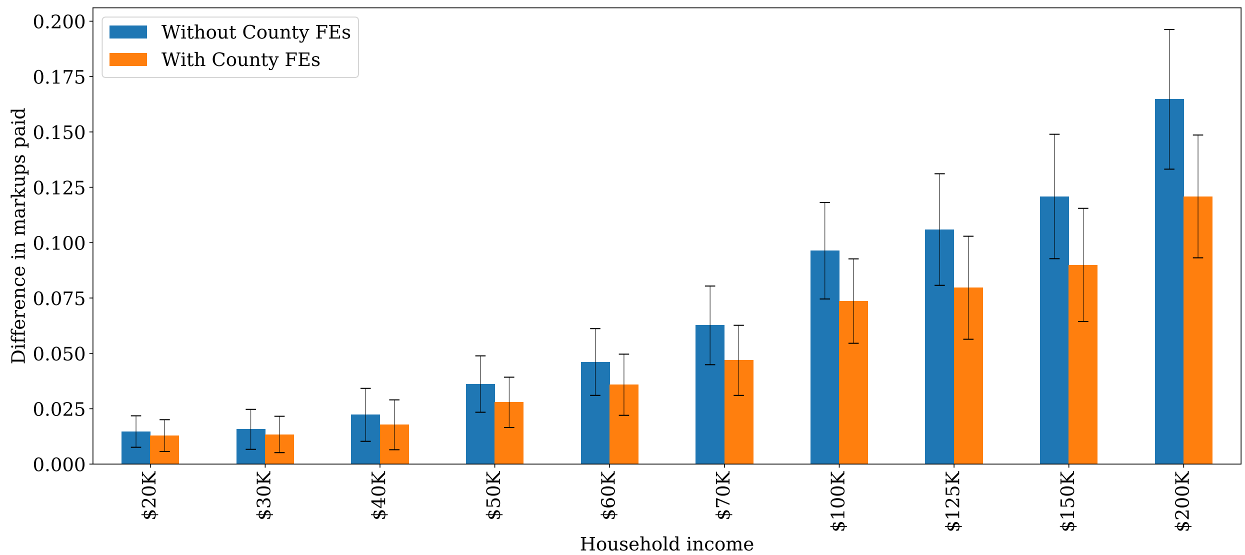
Differential volume discounts by retailer size. While the store fixed effects in specification (2) absorb systematic differences in wholesale costs, transport costs, and local input costs that cause marginal costs to differ by store, they will not absorb heterogeneity in marginal costs by product-store pair. This is problematic if some retailers face lower marginal costs on a subset of items. For example, suppose large retailers are able to negotiate volume discounts and that these volume discounts are steeper on a subset of products (e.g., on commodity items compared to luxury items). In this case, our data would overstate the marginal cost and understate the markup on commodity items sold at large retailers. If low-income households buy more commodity items at large retailers than high-income households, this would lead us to overestimate the difference in markups paid across income groups.

I address this concern by testing whether the difference in markups paid by low- and high-income households is driven by large retailers in the sample. I rank retailers by total sales in the Homescan data. Then, I test specification (2) for the subsample of transactions excluding the largest retailer, excluding the largest three retailers, excluding the largest five retailers, and so on.

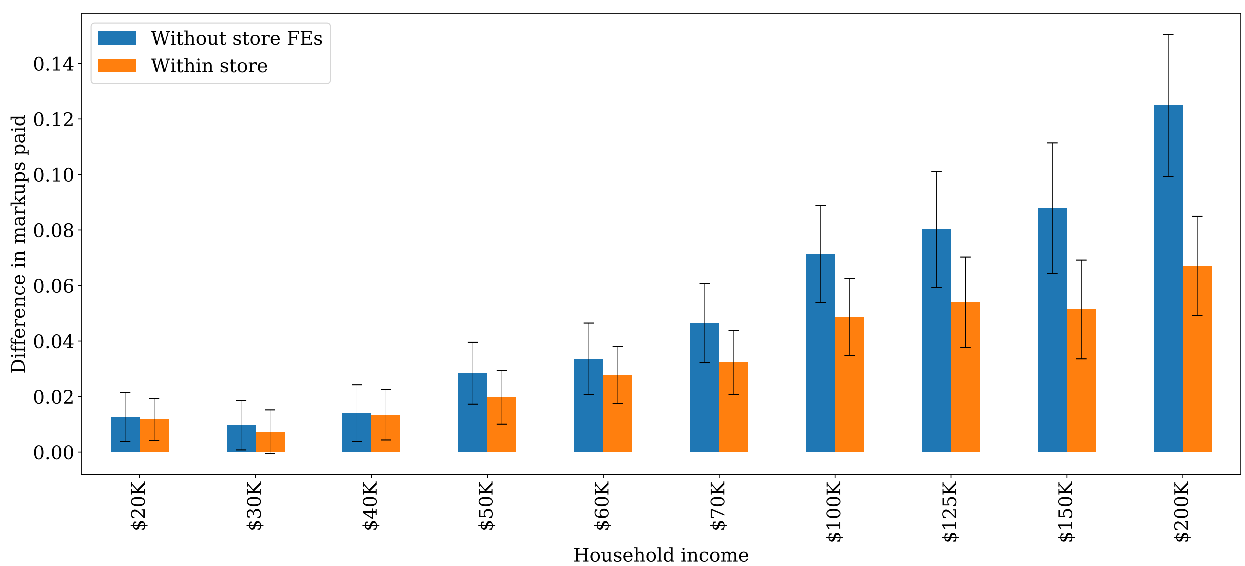
[21] While I use the retail markup in levels as the dependent variable for all regression results in the main text, I replicate all analyses using the log retail markup as the dependent variable in Appendix Figure B6. The results are similar.

[22] A review of the data suggests that Nielsen store IDs are only available for the subset of retailers that also participate in Nielsen's retailer scanner data program. One reason the difference in markups paid between low- and high-income groups is smaller in the subset of transactions made at Nielsen-identified stores could be that these retailers are more homogeneous than retailers in the full sample.

Figure 2: Difference in markups paid relative to households with below $20K income.



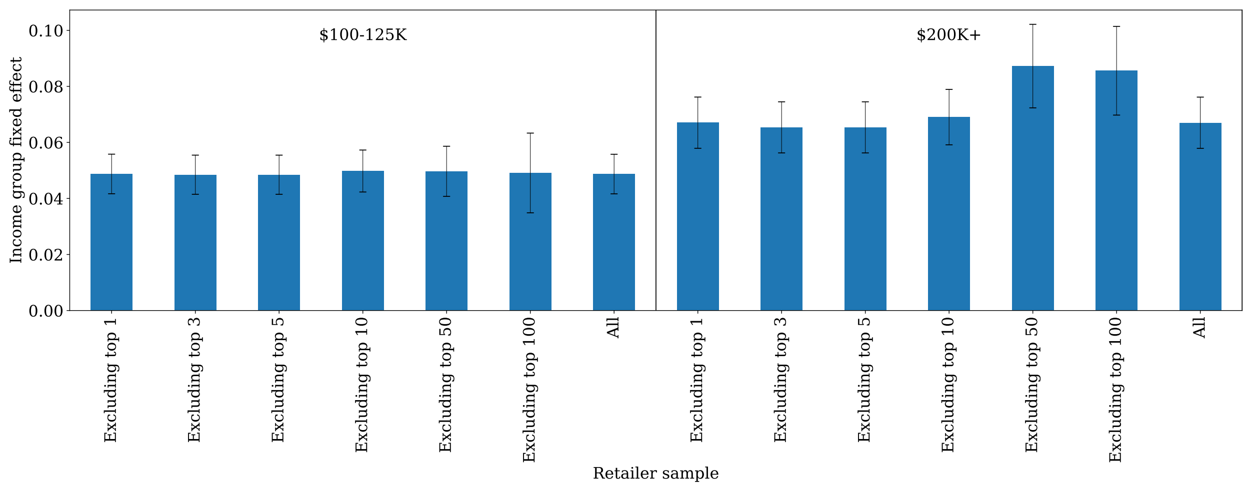
(a) With and without county fixed effects (N = 25.8 million).



(b) With and without store fixed effects (store transactions only, N = 14.0 million).

Note: These figures plot the coefficients β_ℓ on household income dummies in a sales-weighted regression of the markup paid by a household on demographic controls (race, ethnicity, household size, and presence and age of female head of household). Income levels on the y-axis are the minimum of the income bracket provided by Nielsen (e.g., $30K includes households reporting income between $30–40K). Standard errors are two-way clustered by product brand and household county. Figure (a) shows β_ℓ with and without county fixed effects (specification (1)), and (b) shows $\tilde{\beta}_\ell$ with and without store fixed effects for the sample of transactions where the unique store ID is observed (specification (2)).

Figure 3: Stability of markup gap excluding largest retail chains.



If mismeasurement of marginal costs at large retailers drives the effect, then the coefficients on income should diminish as large retailers are removed from the sample.

Figure 3 plots the fixed effect coefficient on two of the highest income groups as I successively remove large retailers from the sample. The coefficient on income is stable, suggesting that the observed markup gap is not driven by differential quantity discounts at large retailers.

Comparison to differences in prices paid for identical products. Previous work by Aguiar and Hurst (2007a), Broda et al. (2009), and Pisano et al. (2022) documents that high-income households tend to pay higher unit prices for the same UPCs as low-income households. If wholesale costs for each UPC are constant across retailers and over time, differences in prices paid identify differences in retail markups. How much of the markup gap between low- and high-income households presented here is due to differences in prices paid for identical products?

To answer this question, I consider specifications (3) and (4):

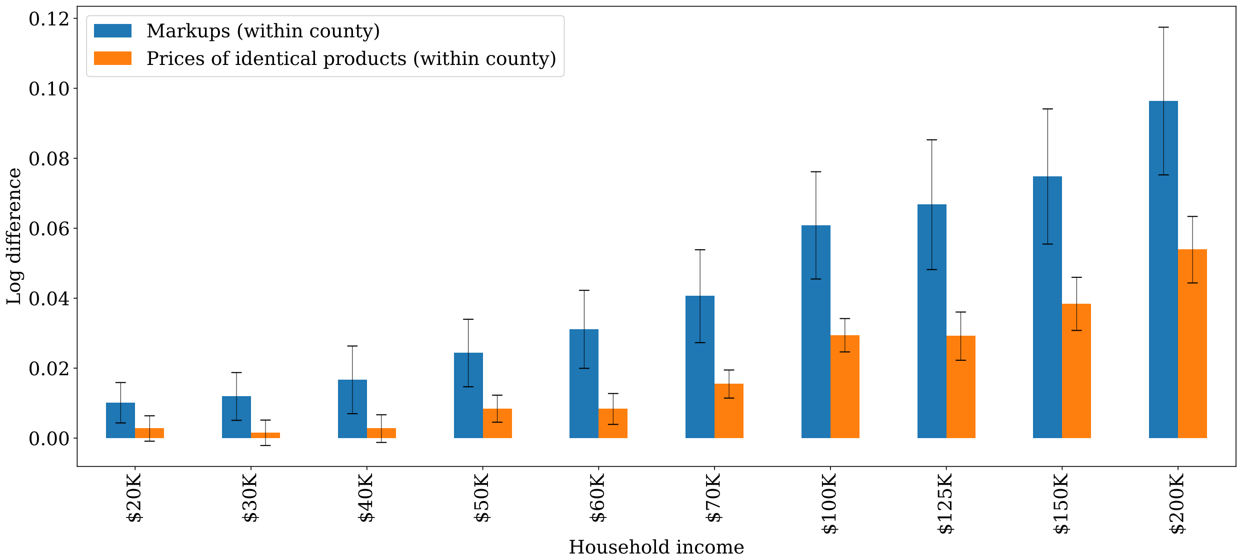
$$\log \text{Markup}_{i,g,t} = \sum_{\ell} \beta_\ell 1\{i \text{ has income level } \ell\} + \gamma' X_i + \delta_{\text{County}} + \epsilon_{i,g,t}, \tag{3}$$

$$\log \text{Price}_{i,g,t} = \sum_{\ell} \beta_\ell 1\{i \text{ has income level } \ell\} + \gamma' X_i + \kappa_g + \delta_{\text{County}} + \epsilon_{i,g,t}, \tag{4}$$

where κ_g are UPC fixed effects. Since retail markups are $\text{Markup}_{i,g,t} = \text{Price}_{i,g,t}/\text{Wholesale cost}_{g,t}$, the coefficients β_ℓ measured under each specification will be identical if costs for each good are constant and if all households have identical expenditure shares across goods.

Figure 4 plots the fixed effects on income level in specifications (3) and (4). The gap in log markups is twice as large as the gap in log prices paid for identical products. Appendix Figure B1 shows that the markup gap is greater because of differences in basket composition across households in different income groups: high-income households also tend to substitute to

Figure 4: Difference in log markups paid within county (blue) and difference in log prices paid for identical products within county (orange) relative to households with below $20K income.



Note: The blue bars plot the coefficients β_ℓ from a sales-weighted regression (specification (3)) of log markup paid on household income dummies, demographic controls, and county fixed effects. The orange bars plot the coefficients β_ℓ from an analogous regression (4) of log price paid on household income dummies, demographic controls, county fixed effects, and UPC fixed effects. Standard errors are two-way clustered by product brand and household county.

products with higher average markups.[23] The use of wholesale cost data provides common units under which different products can be compared.

Elasticity of markups to household income. An alternative way to measure the link between retail markups paid and income is to estimate the elasticity of markups to household income. This is the approach taken by Broda et al. (2009), who estimate the elasticity of prices paid for identical products to household income in Nielsen data from 2005.

Table 1 reports estimates of the elasticity of retail markups to household income.[24] The estimated elasticity is 3.1% (column 1) and 3.8% after controlling for demographic characteristics (column 2). The parallel estimates in Broda et al. (2009) are 1.1% and 1.3%.[25] Hence, the elasticity of markups to household income is about two times greater than the elasticity of prices paid for

[23]In fact, the gap in markups paid *for identical products* is slightly smaller than the gap in prices paid for identical products (see Appendix Figure B2), since reductions in the retail price of a UPC often coincide with reductions in its cost to retailers.

[24]As noted by Broda et al. (2009), Nielsen reports income in discrete categories, and thus a continuous measure of household income is not available. I follow Broda et al. (2009) and recode each household's income as the midpoint of the income bracket. For example, a household earning $13,000 is part of the $12,000-$15,000 income group and is assigned an income equal to $13,500. For the group with over $200,000 in annual income, I assign an income of $225,000.

[25]Appendix Table B5 replicates the analysis from Broda et al. (2009) for the 2007 Homescan panel and yields estimates in the range 1.5-2.0%, slightly larger than those reported by Broda et al. (2009). These differences are due to the panel year used, minor differences in the demographic controls included (Broda et al. 2009 also include marital status and use city, rather than county, controls), and sales-weighting.

identical products to income.

The elasticity of markups paid to household income within county (3.0%) and within store (2.0%) account for about three-quarters and about half of the overall elasticity of markups to income. Hence, the portion of the link between markups and income that persists within county and within store is similar when measured this way.

Decomposing differences within store. The remaining columns 6–8 of Table 1 explore the elasticity of markups to household income within store-product group, within store-product module, and within store-UPC. Two-thirds of the link between markups and household income at the store level is attributed to differences in markups paid within product modules. At the finest level of disaggregation, differences in markups paid for the same UPC at the same store constitute about 40 percent of the within-store elasticity and one-fourth of the unconditional elasticity. Similar results obtain if I estimate fixed effects on income level in a regression of markups on income and store, store-product module, and store-UPC fixed effects (see Appendix Figure B3).[26]

Robustness. In the main text, I use a national wholesale cost for each UPC in each month, use base prices from PromoData as the measure of wholesale costs, and use sales-weighted regressions. Appendix Table B1 shows that the results are not sensitive to these choices: the share of the markup gap that persists within county and within store is similar if I use only transactions matched to wholesale costs in the same market, if I use PromoData deal prices as the measure of retailers' wholesale costs, and if I instead use a cost-weighted specification or use log retail markups as the dependent variable.[27] Appendix Table B1 also reports results using only the hundred most frequently purchased product modules, excluding demographic controls, excluding perishable items, and for years 2006, 2008, and 2009. Results from all these alternatives are quite similar.

3.2 Fact 2. Products with high-income consumers have higher markups

The previous section showed that high-income households pay higher markups. In principle, the gap in markups paid between low- and high-income households could be driven by high-income households paying higher markups on each product, with no systematic correlation between a product's average markup and the income of its buyers. In this section, I show that products bought by high-income consumers tend to have higher retail markups. I also show that the link between products' retail markups and buyer income is not driven by other supply-side factors.

[26]Differences in markups paid by store-UPC result from coupon usage and temporal variation in prices due to sales. Appendix Figure B4 shows that savings from coupon usage are non-monotonic in income (echoing findings by Griffith et al. 2009), which suggests that a large part of differences in markups paid within store-UPC result from differences in the propensity to exploit temporal variation in prices.

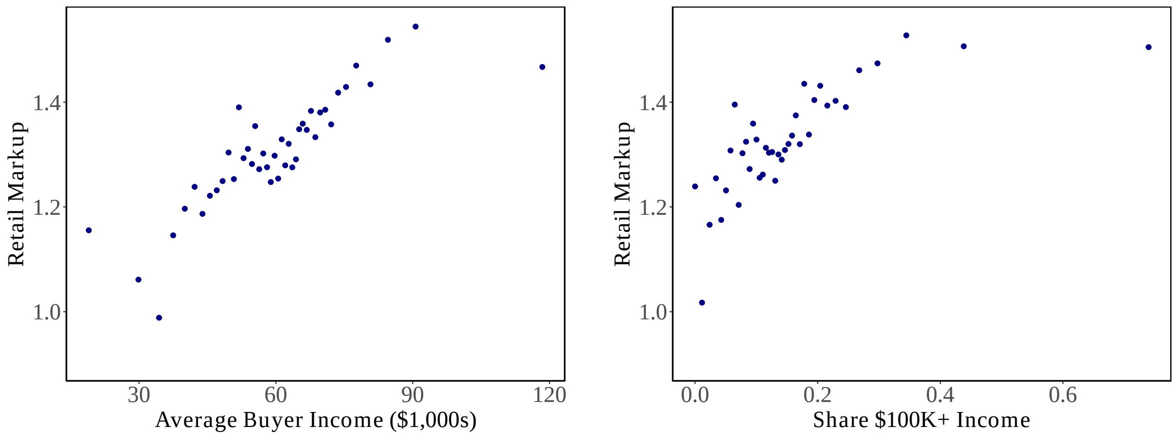
[27]Appendix Figures B5, B6, B7, and B8 explicitly replicate Figure 2 using market-matched wholesale costs, log markups, cost-weighted regressions, and PromoData deal prices. Appendix Tables B2 and B3 also show that the elasticity of markups to household income is similar using market-matched wholesale costs or PromoData deal prices as a measure of wholesale costs.

Table 1: Impact of income on markups paid.

Log Retail Markup	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Household Income	0.031**	0.038**	0.030**	0.025**	0.020**	0.016**	0.015**	0.009**
	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Household Size		−0.007**	−0.006**	−0.007**	−0.006**	−0.005**	−0.005**	−0.003**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Demographic Controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FEs			Yes	Yes	Yes	Yes	Yes	Yes
Store FEs					Yes	Yes	Yes	Yes
Store-Group FEs						Yes		
Store-Module FEs							Yes	
Store-UPC FEs								Yes
N (millions)	25.8	25.8	25.8	14.0	14.0	14.0	14.0	14.0
R^2	0.00	0.00	0.02	0.02	0.08	0.42	0.63	0.92

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $13.5K). Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

Figure 5: UPC retail markup and buyer income.



Note: The unit of observation is a UPC. The UPC retail markup is calculated as the sales-weighted average markup over all observed transactions. Graphs show a binned scatter weighted by UPC sales.

To start, I calculate the average markup for each product barcode (UPC) in the Homescan data as the sales-weighted average over all observed transactions of the product. Figure 5 shows binned scatterplots of these UPC retail markups against two measures of buyer income. In both cases, retail markups appear positively correlated buyer income.

Table 2 tests this relationship. Consistent with the visual correlation, columns 1, 3, and 5 show that a UPC's retail markup is positively associated with average buyer income and the share of high-income buyers and negatively associated with the share of low-income buyers. A 10pp increase in the share of purchases coming from households with over $100K in income is associated with a 7 percent higher retail markup.

While these associations are statistically significant, the low *R*-squared of these regressions suggests there is substantial variation in retail markups across UPCs driven by other factors. Hence, columns 2, 4, and 6 add product module fixed effects. All three measures of buyer income remain significantly associated with UPC retail markups, though the magnitudes shrink by half. Accordingly, for two products in the same product module, a 10pp increase in the share of purchases coming from households with over $100K in income is associated with a 3 percent increase in the product's retail markup.[28]

Supply-side factors. In representative household models, firms may charge heterogeneous markups due to differences in the representative household's elasticity of demand across goods, differences in market power across firms (as in the nested oligopoly model of Atkeson and Burstein

[28]While product module fixed effects absorb factors that may cause retail markups to differ across product categories, some differences in markups across product modules may be due to correlation with buyer income. For example, Appendix Figure B10 shows patterns of substitution across substitutable product categories (e.g., butter and margarine, and potato chips and tortilla chips) correlated with income. In those examples, the product modules with higher-income customers have higher markups.

Table 2: Relationship between buyer income and UPC retail markup.

Log UPC Retail Markup	(1)	(2)	(3)	(4)	(5)	(6)
Log Avg. Income	0.289**	0.100**				
	(0.063)	(0.038)				
Share $100K+ income			0.686**	0.299**		
			(0.133)	(0.067)		
Share $10-50K income					−0.357**	−0.101**
					(0.079)	(0.044)
Product Module FEs		Yes		Yes		Yes
N	67 161	67 161	67 161	67 161	67 161	67 161
R^2	0.03	0.47	0.03	0.47	0.02	0.47

Note: The unit of observation is a UPC. The dependent variable is the UPC retail markup, which is the sales-weighted average of markups at which the UPC was sold. Regression weighted by UPC sales, and standard errors clustered by product brand. ** indicates significance at 5%.

2008), and/or differences in the extent of variety available (e.g., Hummels and Lugovskyy 2009). In Table 3, I test whether the association between retail markups and buyer income is driven by these factors.

Column 1 of Table 3 reports the elasticity of the UPC retail markup to average buyer income before adding any additional controls (this is identical to column 1 in Table 2). In column 2, I add a proxy for whether a product module is a necessity. This measure of necessity is the share of households in the panel that purchase any product from the module in the year. Product modules that rank high on this necessity index include fresh bakery bread, refrigerated dairy milk, and toilet tissue, while hair nets and caps, dishwashing accessories, and hair accessories rank low on this measure. Surprisingly, this proxy of necessity appears (mildly) negatively correlated with UPC markups, and the coefficient on buyer income is similar.[29]

Models such as Hummels and Lugovskyy (2009) suggest that the extent of product variety and whether consumers have "finicky" tastes can affects firms' elasticities of demand and thus markups. Columns 3 and 4 add the log number of unique brands and unique UPCs in a module, as proxies for the degree of product variety. These measures of the product variety are negatively correlated with retail markups, but the estimated coefficient on buyer income does not meaningfully change.

Column 5 adds the sales share of the firm selling the UPC in the product module and the Herfindahl-Hirschman index of sales concentration within the product module.[30] Neither measure appears to significantly covary with UPC markups.[31] Column 6 shows that UPC sales shares are

[29]This finding contrasts with models that generate heterogeneous markups by bounding marginal utility for a product from above, such as Simonovska (2015) and Neiman and Vavra (2019). Those models typically predict that "cutoff" products consumed by the fewest households have low, rather than high, markups.

[30]The firm owning the UPC is identified using data from Global Standards One (GS1), which issues product barcodes to firms. See Appendix A for more detail.

[31]These results suggest the degree of firm ownership is not a primary determinant of retail markups. In fact, in the

Table 3: Relationship between buyer income and UPC retail markup.

Log UPC Retail Markup	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log Avg. Income	0.289**	0.301**	0.289**	0.280**	0.293**	0.286**	0.097**	0.097**	0.101**	0.104**	0.112**	0.100**
	(0.063)	(0.062)	(0.062)	(0.059)	(0.063)	(0.063)	(0.035)	(0.036)	(0.035)	(0.035)	(0.031)	(0.036)
Necessity		−0.054						−0.149**				
		(0.038)						(0.024)				
Log No. Unique Brands			−0.018**						−0.014**			
			(0.006)						(0.007)			
Log No. Unique UPCs				−0.040**						−0.026**		
				(0.007)						(0.005)		
Firm Sales Share					−0.028						−0.088**	
					(0.073)						(0.031)	
Module HHI					0.019						0.144**	
					(0.076)						(0.060)	
UPC Sales Share						0.351**						0.170*
						(0.178)						(0.099)
Brand Sales Share						0.078						−0.015
						(0.057)						(0.033)
Product Group FEs							Yes	Yes	Yes	Yes	Yes	Yes
N	67 161	67 161	67 161	67 161	67 161	67 161	67 161	67 161	67 161	67 161	67 161	67 161
R^2	0.03	0.03	0.03	0.06	0.03	0.04	0.30	0.30	0.30	0.30	0.30	0.30

Note: The unit of observation is a UPC. The dependent variable is the UPC retail markup, which is the sales-weighted average of markups at which the UPC was sold. Regression weighted by UPC sales, and standard errors clustered by product brand. Necessity is measured as the share of households in the Homescan panel that purchase at least one good from that module in the year. UPC sales share, brand sales share, firm sales share, and the product module Herfindahl-Hirschman index are all calculated using sales in the Homescan data. * indicates significance at 10%, ** at 5%.

positively correlated with retail markup, as the Atkeson and Burstein (2008) oligopoly model would suggest, but the coefficient on buyer income is unchanged.

Columns 7–12 replicate this analysis after adding product group fixed effects. Across these regressions, the estimated elasticity of UPC retail markup to buyer income within-product group is stable around 10 percent. While this is not an exhaustive exploration of all potential supply-side mechanisms, it suggests that the link between buyer income and retail markups is not driven by a spurious correlation with supply-side factors.

Robustness. Appendix Tables B6 and B7 estimate the relationship between retail markups and buyer income in the panel (rather than projecting on the cross-section of products as in Table 3) and control for the same supply-side factors as in Table 3. They yield similar conclusions: the association between retail markups and buyer income does not meaningfully change after controlling for supply-side characteristics.

3.3 Fact 3. Markups depend positively on other buyers' incomes

Finally, I explore how markups paid by a household depend on the income of other buyers. Measuring this cross-elasticity is important to understand how the macro elasticity of markups to income—the rate of change in the economy's aggregate markup as all households' incomes change—compares to the micro elasticity of markups to income observed in the cross-section. To a first order, the macro elasticity is given by the sum of the micro elasticity of markups to own income and the cross-elasticity of markups to others' incomes:

$$\varepsilon^{\text{agg}}_{\mu,\text{Income}} \approx \varepsilon^{\text{individual}}_{\mu,\text{Own Income}} + \varepsilon^{\text{individual}}_{\mu,\text{Others' Incomes}}.$$

Suppose retailers are able to perfectly price discriminate across customers, so that the markup paid by a household depends only on its own characteristics, and not the characteristics of other buyers. In that case, the cross-elasticity $\varepsilon^{\text{individual}}_{\mu,\text{Others' Incomes}} = 0$, and the macro elasticity of markups to income is exactly equal to the micro elasticity. On the other hand, imperfect price discrimination or the presence of other spillovers means that the macro and micro elasticities can differ.[32]

I first show evidence in the cross-section that the cross-elasticity $\varepsilon^{\text{individual}}_{\mu,\text{Others' Incomes}}$ is positive and large. I then move to the time series, and show that the cross-elasticity measured in the time series appears consistent with that measured in the cross-section.

Cross-sectional evidence. Figure 6 shows the average retail markup paid by households in five income groups split by (a) quintile of county income and (b) quintile of the average income of a

data, two products owned by the same firm in the same product category often have substantially different markups. Within firm × product module, the elasticity of a UPC's retail markup to buyer income is 14.5%.

[32] As I discuss at the close of this section, the cross-elasticity may be positive or negative in theory. For example, in Varian (1980), an increase in the share of uninformed buyers (a proxy for the share of price-insensitive high-income buyers) decreases the average markup paid by informed buyers, due to intensifying competition. The model developed in Section 4 of this paper instead generates a positive cross-elasticity, consistent with the empirical evidence.

Table 4: Impact of own and other buyers' incomes on markups paid.

Log Retail Markup	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) IV
Log Household Income	0.038**	0.033**	0.053**	0.016**	0.024**
	(0.004)	(0.004)	(0.008)	(0.001)	(0.003)
Log Avg. CBSA Income		0.104**	0.092**		
		(0.010)	(0.011)		
Log Avg. Income: Other UPC Buyers				0.074**	0.073**
				(0.028)	(0.028)
Demographic Controls	Yes	Yes	Yes	Yes	Yes
County FEs				Yes	Yes
Store FEs				Yes	Yes
Product Module FEs				Yes	Yes
N (millions)	25.8	23.8	23.8	14.0	14.0
R^2	0.00	0.01	0.01	0.35	0.35

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen. In columns 3 and 5, log household income is instrumented using the education and occupation of the male and female heads of household. Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

UPC's other buyers. Consistent with Section 3.1, the average markup paid by the highest income group lies above the average markup paid by the lowest-income group in all cases. Conditional on income, however, markups paid also increase with county income and with the income of other buyers of the same product.

Table 4 tests how a household's own income and the income of other buyers affects the markup paid by the household. The elasticity of markups paid to the average CBSA income in column 2 is positive and nearly three times larger than the elasticity of markups paid to own income (10.4% compared to 3.3%). Since variation in unobserved components of marginal cost across CBSAs may bias this estimate, in column 4 I instead use the average income of other buyers of the same UPC as a measure of other buyers' income, and include county and store fixed effects to control for unobserved components of marginal cost that vary at the county and store level and product module fixed effects to control for variation in price sensitivity by product category. A similar result obtains: the elasticity of markups paid to the income of other buyers of the same UPC is large and positive at 7.4%.

One concern is that noise in reported household income attenuate the elasticity of markups to own household income and bias upward the estimated cross-elasticities of markups to other buyers' incomes. Columns 3 and 5 instrument for household income using the education level and occupation group of both the male and female heads of household. As expected, the estimated elasticity of markups to own income increases, and the estimated elasticity of markups to other

Table 5: Relationship between CBSA average markup, income level, and inequality.

	Log Sales-Weighted Avg. Markup				*Log Cost-Weighted Avg. Markup*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Average CBSA Income	0.131**	0.107**	0.110**	0.120**	0.112**	0.094**	0.096**	0.104**
	(0.019)	(0.013)	(0.014)	(0.015)	(0.015)	(0.012)	(0.013)	(0.014)
Top 10% Income Share		0.157**				0.118**		
		(0.058)				(0.045)		
Top 5% Income Share			0.113*				0.083	
			(0.068)				(0.055)	
Top 1% Income Share				0.068				0.048
				(0.087)				(0.071)
N	882	881	881	881	882	881	881	881
R^2	0.34	0.36	0.35	0.34	0.30	0.32	0.31	0.30

Note: The unit of observation is a CBSA. The dependent variable is the average sales-weighted markup (columns 1-4) or cost-weighted markup (columns 5-8) for purchases observed in the CBSA. CBSA Income is from the BEA, and CBSA inequality data is for 2013 from Sommeiller et al. (2016). Regressions are weighted by CBSA sales, with robust standard errors. * indicates significance at 10%, * at 5%.

buyers' incomes slightly attenuates. Nevertheless, the elasticity of markups paid to CBSA income (9.2% in column 3) and other UPC buyers' incomes (7.3% in column 5) remains large and positive.

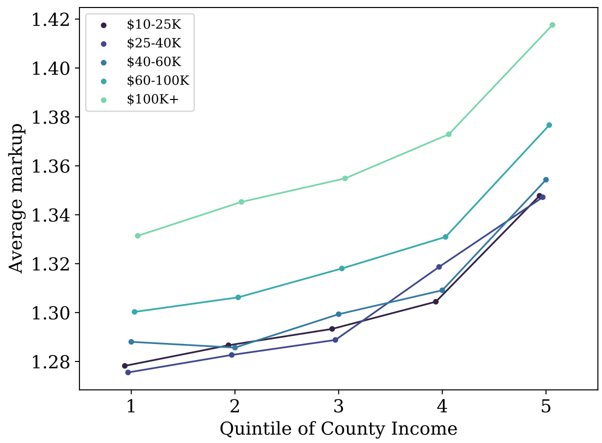
Adding the elasticity of markups to own income and to others' incomes results in an estimated macro elasticity $\varepsilon^{\text{agg}}_{\mu,\text{Income}}$ of 8–15%. For comparison, Figure 7 plots average retail markups in each CBSA against average CBSA income. The best-fit line has an elasticity of 13.1%, squarely in the range of macro elasticities estimated in the panel data. Table 5 further tests how these average retail markups across CBSAs covary with income and with the measures of CBSA inequality. The estimated macro elasticity of markups to income is between 9–13%, and higher inequality (though inequality at the very top end less so) is also positively associated with CBSA markups.

Time series evidence. Up to this point, all analyses have exploited cross-sectional variation using a single year of data. A natural question is to what extent this positive dependence on other buyers' incomes persists in the time series.

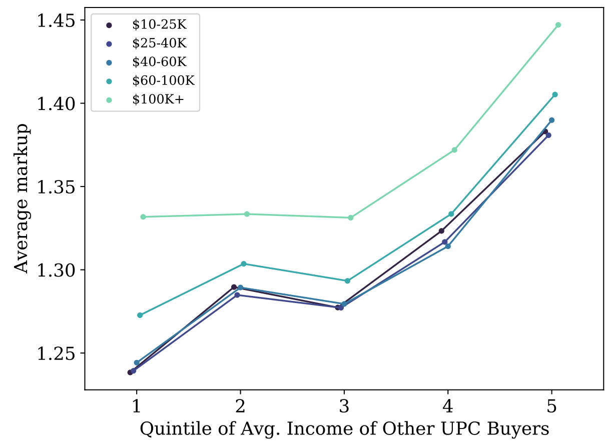
To answer this question, I undertake two analyses. First, I use retail markups for the years 2006–2012, which is the period over which the PromoData wholesale cost data are available. Second, I use retail prices over a longer time period (2005–2019) for which Nielsen Homescan data is available. Both analyses suggest that the positive dependence of markups to other buyers' incomes extends to the time series and that macro elasticities measured in the cross-section and time series are similar. I discuss each in turn.

For each year from 2006–2012, I construct retail markups using the Nielsen Homescan data and PromoData wholesale cost data in the same fashion as described in Section 2. Since the set of UPCs covered by PromoData vary substantially from year to year, in Table 6 I test how within-UPC changes in buyer income affect markups paid by households. Within UPC, markups paid by a

Figure 6: Average retail markup paid by income group, split by (a) quintile of county income and (b) average income of other UPC buyers.

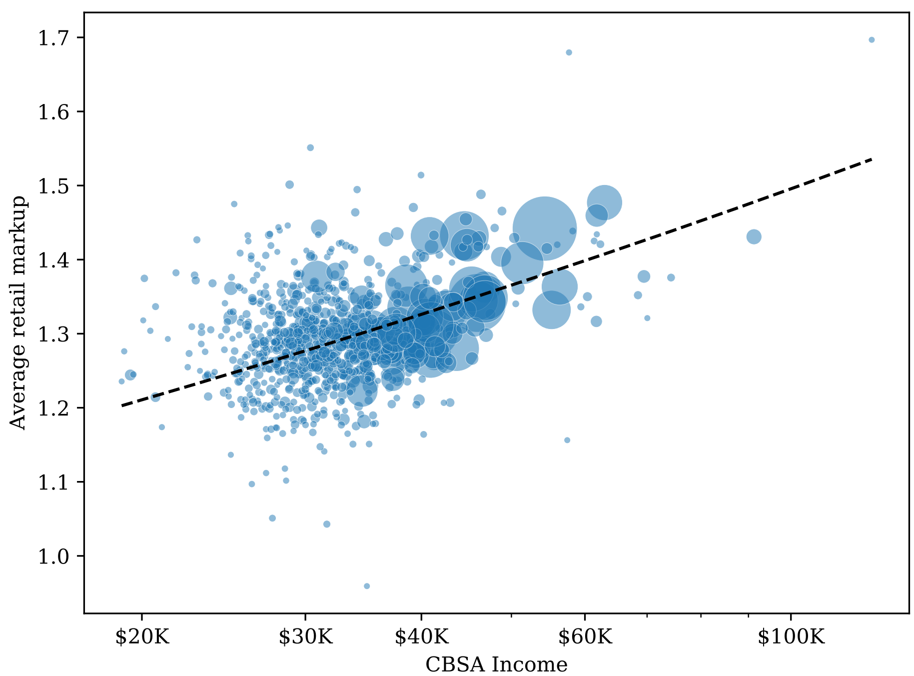


(a) County income



(b) Income of UPC's other buyers

Figure 7: CBSA per-capita income and average retail markup.



Note: Each bubble refers to a CBSA, and the size of the bubble is proportional to total CBSA sales in the Nielsen data. Average CBSA income in 2007 is from the BEA. The dotted curve plots the best fit from a regression on log markup on log CBSA income weighted by CBSA sales.

Table 6: Impact of own and other buyers' incomes on markups over time, 2006–2012.

Log Retail Markup	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS
Log Household Income	0.018**	0.023**	0.017**	0.017**	
	(0.001)	(0.002)	(0.001)	(0.002)	
Log Avg. Income: Other UPC Buyers in Year	0.102**	1.085**			
	(0.025)	(0.275)			
Log Avg. Income: Other Brand Buyers in Year			0.449**	1.482**	
			(0.131)	(0.340)	
Log Avg. Income: All UPC Buyers in Year					0.139**
					(0.028)
UPC FEs	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	
N (millions)	144	128	144	128	144
R^2	0.55	0.53	0.55	0.56	0.54

Note: Sample includes Homescan purchases with PromoData wholesale costs from 2006–2012. All variables are deflated to 2007 USD using the GDP deflator from FRED. Standard errors two-way clustered by brand and county. Columns 2 and 4 instrument for the average income of other UPC and brand buyers using a shift-share instrument that exploits changes in average income by retailer over time (UPCs that are never purchased at retailers with unique IDs are excluded from the regression). ** indicates significance at 5%.

household depend positively on the income of the UPC's and brand's other buyers in the year of purchase. IV estimates of these spillovers effects (which address endogeneity problems arising from, say, firms choosing which customer segments to target over time) are also positive, though much larger.[33] While the magnitude of the cross-elasticity varies across columns 1–4, in all cases it is similar to or greater than the cross-elasticity documented in the cross-section. In column 5, a reduced form estimate of the elasticity of a UPC's markup to buyer income over time is 13.9%, squarely in the range of macro elasticities estimated in the cross-section.

Since this analysis is constrained by data availability to a short time series, I use the longer time series of prices paid by Nielsen Homescan panelists to extend the analysis. These data cover all products purchased by Homescan panelists in Nielsen-tracked categories. Hence, I no longer include UPC fixed effects (which controlled for changes in PromoData coverage in the previous analysis), but use product module fixed effects since wholesale cost data are not available to facilitate comparisons across products sold in different units.[34]

Table 7 reports the elasticity of log unit prices within a product module to household income and average CBSA and county income. Column 1 reports an elasticity across both time and space

[33]Columns 2 and 4 instrument for the average income of other buyers using the shift-share instrument $\widehat{x}_{g,t} = \sum_r \bar{s}_{r,g}(\text{Average Buyer Income}_{r,t})$, where $\bar{s}_{r,g}$ denote the share of sales of UPC g at retailer r over all years in the sample and Average Buyer Income$_{r,t}$ is the average income of all buyers at retailer r in year t.

[34]Some product modules contain items with more than one unit of account (e.g., some butter products are reported in ounces while others are reported in pounds). Where this is the case, I further disaggregate product modules into groups of products in the same product module measured in the same unit of account.

to CBSA income of 17.6%. Columns 2–5 disaggregate this into an elasticity across space and an elasticity over time. In column 2, I add year-product module fixed effects to isolate how cross-sectional variation in CBSA income affects log unit prices. The estimated elasticity is 16.0%. In column 3, I use CBSA-product module fixed effects to isolate how variation in CBSA income over time affects log unit prices. The time-series elasticity estimate is larger than the cross-sectional estimate, at 32.6%. One concern is that changes in income over time may be confounded by other changes over time; hence columns 4 and 5 use two-way fixed effects to estimate cross-sectional and time-series elasticities. Column 4 includes year-product module fixed effects and CBSA fixed effects, so that the estimate corresponds to the cross-sectional elasticity of log unit prices to deviations in CBSA income. The estimated elasticity is 10.5%. Column 5 instead includes CBSA-product module fixed effects and year fixed effects, so that the estimate corresponds to the elasticity of deviations in log unit prices in each year to changes in CBSA income over time. The resulting elasticity estimate is 10.2%. Columns 6–10 repeat this analysis instead using county income and find similar results: the time series elasticity of log unit prices to CBSA or county income is similar or larger in magnitude than the cross-sectional elasticity.[35]

Discussion. The empirical evidence from both the cross-section and the time series places the macro elasticity of markups to income around 8–15%, larger than the 3–4% observed in the cross-section. This estimate is broadly in line with previous evidence from the trade literature. For example, Simonovska (2015) finds that the elasticity of markups to per-capita income of the export destination is 12–24%. Time series evidence in the U.S. also accords with these findings. Stroebel and Vavra (2019) report an elasticity of retail markups to housing wealth of 15–20%. Most recently, Bhardwaj et al. (2022) compare how prices evolve in counties that win and lose bids for million dollar plants and find an elasticity of retail prices to changes in hourly wages of about 10%.

This empirical evidence also relates to a literature that documents differences in price indices across space, such as Handbury (2021) and Diamond and Moretti (2021), and differences in the evolution of price indices over time, such as Jaravel (2019). That literature finds that prices and markups paid by households depend substantially on the characteristics of other buyers. Note, however, that the net direction of these spillovers is ambiguous ex ante. For example, pro-competitive effects of firm entry may lead to declines in prices and markups paid for goods consumed by high-income households as the share of high-income households rises. (In line with such pro-competitive effects, Jaravel (2019) finds that product categories consumed disproportionately by high-income households experience increases in product variety and lower inflation as the share of high-income households grows, and Handbury (2021) finds that products consumed by high-income households are relatively less expensive in high-income areas.) On the other

[35] One might be concerned that, since the sample years include the Great Recession, the time series elasticity measured here reflects changes in income over the business cycle rather secular trends in income over time. To address this concern, Appendix Table B8 reports estimates of the elasticity of log unit prices to CBSA and county income using only 2004 and 2019. While the statistical power is reduced, the relative magnitudes of the cross-section and time series elasticities are qualitatively similar.

Table 7: Elasticity of unit prices to CBSA and county income across space and over time, 2005–2019.

Log Unit Price	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Household Income	0.083**	0.084**	0.081**	0.082**	0.082**	0.083**	0.084**	0.079**	0.080**	0.079**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log Avg. CBSA Income	0.176**	0.160**	0.326**	0.105**	0.102**					
	(0.012)	(0.009)	(0.089)	(0.014)	(0.009)					
Log Avg. County Income						0.148**	0.135**	0.309**	0.091**	0.093**
						(0.009)	(0.007)	(0.083)	(0.014)	(0.015)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Module FEs	Yes					Yes				
Year-Product Module FEs		Yes					Yes			
CBSA-Product Module FEs			Yes							
Year-Product Module FEs and CBSA FEs				Yes						
CBSA-Product Module FEs and Year FEs					Yes					
County-Product Module FEs								Yes		
Year-Product Module FEs and County FEs									Yes	
County-Product Module FEs and Year FEs										Yes
N (millions)	392	392	392	392	392	418	418	418	418	418
R^2	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.88	0.87	0.88

Note: Sample includes all Homescan purchases in odd years from 2005–2019. For the purpose of this analysis, "Product Modules" are defined as product modules from Nielsen further disaggregated by unit of account. For example, all butter products with units in ounces are counted as a module, and distinct from butter products with units in pounds. All unit prices, incomes, and sales are deflated to 2007 USD using the GDP deflator from FRED, and regressions are weighted by (deflated) sales. Standard errors two-way clustered by brand and county. * indicates significance at 10%, ** at 5%.

hand, frictions that limit retailers from perfectly price discriminating across customers (such as the uniform retail pricing documented by DellaVigna and Gentzkow 2019) lead to positive spillovers across households. For retail markups, the data presented in this section suggest the net effect is a positive dependence of markups on other buyers' incomes.

4 A Search Model of Income and Markups

In this section, I develop a model in which differences in markups paid by households arise endogenously due to differences in search effort. This micro-foundation draws on a large literature that documents the importance of search effort in prices paid (e.g., Aguiar and Hurst 2007a, Alessandria and Kaboski 2011, Kaplan and Menzio 2015). The model is able to account for the three pieces of evidence documented in the previous section: (1) markups increase with household income, (2) firms selling to a high-income customer base have higher markups, and (3) the macro elasticity of markups to income is greater than the micro elasticity in the cross-section.

Since the role of the model is to demonstrate how changes in the demand side of the economy affect markups, I begin with a single-good version of the model, in which the supply side of the economy is deliberately simple. Section 4.7 extends the model with heterogeneous products.

While the model rationalizes the empirical evidence presented in the previous section, it is not the only model that can do so. One could consider different search technologies than the one I use here, or different micro-foundations altogether. Online Appendices E and F explicitly develop two alternatives, and I compare these approaches at the end of this section.

4.1 Households

A unit measure of households consume perfectly substitutable varieties of a good sold by a measure M of firms. As in canonical models of search frictions, households know the distribution of prices posted by firms, $F(p)$, but do not know which retailer sells at which price. As a result, households retrieve price quotes from firms before making a purchase.

The number of price quotes observed by household i prior to purchase is a random variable with probability mass function $\{q_{i,n}\}_{n=1}^{\infty}$. That is, with probability $q_{i,1}$, the household observes only a single price quote, with probability $q_{i,2}$, the household observes two price quotes, and so on. Each price quote retrieved is an independent draw from the distribution of prices F.

Upon receiving n price quotes, households compare the minimum price quote received to a reservation price, R, and buy one unit of the good from the firm with the lowest price quote p as long as $p \leq R$. If all n quotes received by the household have a price greater than R, I allow the household to costlessly re-draw n quotes. For each unit of consumption, the household repeats the same search process.[36,37]

[36]The assumption that households search separately for each unit of consumption is made to reflect the observation in Pytka (2018) that shopping for a larger consumption basket takes more time. The relationship between consumption basket size c_i and time spent searching $t(c_i, s_i)$ is made explicit below.

[37]I assume each unit of consumption is infinitesimal, so that integer constraints can be ignored.

The distribution of the number of price quotes received by household i is determined by the household's search intensity, s_i. Formally, the function $\mathcal{S} : s \mapsto \{q_n\}_{n=1}^{\infty}$ maps search intensity to the distribution of price quotes received. Let Q_n be the cumulative mass function associated with $\{q_n\}_{n=1}^{\infty}$. Assumption 1 puts constraints on the function $\mathcal{S}$.

Assumption 1. The search mapping function $\mathcal{S} : s \mapsto \{q_n\}_{n=1}^{\infty}$ is such that:

(A). If $s = 0$, $Q_n = 1$ for all n,
(B). $Q_n(s)$ is weakly decreasing in s for all n and strictly decreasing for $n = 1$,
(C). $Q_n(s)$ is C^{∞} for all n and all $s \geq 0$.

Assumption 1(A) means that if a household spends no time searching, it will always receive exactly one price quote. Under Assumption 1(B), an increase in search intensity leads to a first-order stochastic shift in the number of price quotes received. Assumption 1(C) is made for convenience and is satisfied by common parameterizations of the Burdett and Judd (1983) model.

Households choose search intensity and time spent working to maximize utility subject to a time constraint and budget constraint:

$$\max_{l_i, s_i} \; u(c_i) \qquad \text{s.t.} \qquad \begin{cases} t_i(c_i, s_i) + l_i = 1, & \text{(Time constraint)} \\ p_i c_i = z_i l_i, & \text{(Budget constraint)} \end{cases}$$

where $u(\cdot)$ is a strictly increasing function, c_i is total consumption by household i, $t_i(c_i, s_i)$ is the time it takes household i to shop for c_i units with search intensity s_i, l_i is time spent working with labor productivity z_i, and p_i is the average price paid by i per unit of consumption. With many units of consumption, there is no uncertainty in the average price paid per unit of consumption p_i or the total units of consumption c_i. The budget constraint anticipates that free entry will set firms' profits to zero in equilibrium, so that all household earnings come from labor market work.

I assume that the amount of time it takes to shop for c_i units with search intensity s_i takes the form

$$t_i(c_i, s_i) = \frac{c_i}{a_i} s_i,$$

where a_i is household i's search productivity. That is, time spent shopping increases linearly with the size of the consumption basket c_i, increases linearly with shopping intensity s_i, and decreases with search productivity a_i. Search productivity is allowed to vary across households to capture differences in access to search technologies. For example, access to a car or to a greater density of nearby stores decreases the time required to retrieve a given number of price quotes.[38]

Under utility maximization, households will increase search intensity s_i as long as the marginal

[38] Instead of assuming that time spent shopping is homogeneous of degree one in consumption basket size, one could consider an elasticity of shopping time to basket size that is less than one, reflecting returns to scale in shopping for a larger basket. In practice, differences in search productivity a_i calibrated across households absorb these returns to scale in search.

benefit from increasing search intensity is greater than the marginal cost of doing so. Thus,

$$-\frac{\partial p_i(s_i, F)}{\partial s_i} \leq \phi_i, \tag{5}$$

where the *opportunity cost of search effort* $\phi_i = z_i/a_i$ captures the foregone labor market earnings from increasing search intensity s_i.

If gains from search at any search level are too small—or conversely, if the cost of i's time is too high—a household will choose the corner case $s_i = 0$. For the remainder of the text, I focus on the case where each household has an internal solution for s_i so that (5) holds as an equality.

Aggregate search behavior. Households are indexed by labor productivity $z \in [0, \infty)$, where labor productivity is distributed in the population according to the cumulative density function $H(z)$. I assume there is a one-to-one mapping between labor productivity z and search productivity $a(z)$, so that all households with labor productivity z have the same opportunity cost of search effort $\phi(z) = z/a(z)$. Denote aggregate consumption C and the consumption-weighted distribution of labor productivity,

$$d\Lambda(z) = \frac{c(z)}{C} dH(z).$$

I refer to $\Lambda(z)$ as the *distribution of buyers' incomes* because it captures the cumulative distribution of wages z over the set of purchases.

We can summarize aggregate search behavior using the probability mass function $\{\bar{q}_n\}_{n=1}^{\infty}$, where

$$\bar{q}_n = \int_0^{\infty} q_n(z) d\Lambda(z), \qquad \text{for all } n. \tag{6}$$

This is a probability mass function because $q_n(z)$ sums to one over all n for each z and $d\Lambda(z)$ integrates to one over all values of z. With this description of aggregate search behavior in hand, I proceed to the production side of the economy.

4.2 Firms

A measure M of ex ante identical firms produce output with a constant-returns production technology in labor. I normalize the per-unit variable cost of production to one (i.e., households' labor productivities z are measured relative to the cost of producing one unit of output). Firms set prices to maximize variable profits $\pi(p)$,

$$\max_p \pi(p) = (p - 1)D(p),$$

where the demand curve $D(p)$ that a firm faces depends on its price, the distribution of prices charged by other firms F, and the aggregate search behavior of households.

Following Burdett and Judd (1983), define a *dispersed-price equilibrium* as an equilibrium in which the distribution of prices F is such that all firms choosing a price $p \in \text{supp}(F)$ make identical profits and charging any price $p \notin \text{supp}(F)$ results in strictly lower profits. Proving the existence of a dispersed-price equilibrium follows closely from Burdett and Judd (1983); I relegate the details to Appendix C. Given $\{\bar{q}_n\}_{n=1}^{\infty}$ with $\bar{q}_1 \in (0,1)$, the unique equilibrium price distribution $F(p)$ is

$$F(p) = \begin{cases} 0 & \text{if } p < \underline{p} \\ 1 - \Psi\left[\left(\frac{R-1}{p-1}\right)\bar{q}_1\right] & \text{if } \underline{p} \le p \le R \\ 1 & \text{if } p > R \end{cases} \tag{7}$$

where the lowest price $\underline{p}$ in the support of F is

$$\underline{p} = 1 + \frac{\bar{q}_1}{\sum_{n=1}^{\infty} n\bar{q}_n}(R-1), \tag{8}$$

and $\Psi(\cdot)$ is the inverse of the strictly increasing, C^{∞} function $y(x) = \sum_{n=1}^{\infty} n\bar{q}_n x^{n-1}$.

Firms pay an entry cost of f_e units of labor. Free entry and exit determines the mass of firms M.[39] In equilibrium, the zero profit condition yields

$$\pi(p) - f_e = 0 \qquad \text{for all} \qquad p \in [\underline{p}, R]. \tag{9}$$

4.3 Equilibrium

An equilibrium is a tuple $\left(\{s(z)\}_{z=0}^{\infty}, F, \pi, M\right)$ such that the search intensity $s(z)$ chosen by households with labor productivity z maximizes utility for all z, all firms choosing a price $p \in \text{supp}(F)$ have variable profits π and any price $p \notin \text{supp}(F)$ results in variable profits strictly less than π, the mass of firms M is such that firms make zero profits net of the entry cost, and all resource constraints are satisfied. Equivalently, $s(z)$ satisfies (5) for all z, $F(p)$ is given by (7), and $\pi = f_e$ as in (9).

Define the *aggregate markup* $\bar{\mu}$ as the ratio of total sales to total variable costs. Lemma 1 shows that the fraction of households receiving only one price quote in equilibrium is a sufficient statistic for the aggregate markup.

Lemma 1. *In equilibrium, the aggregate markup is*

$$\bar{\mu} = 1 + (R-1)\,\bar{q}_1.$$

Intuitively, since firms must make identical profits at all prices in the support of F, and the only customers of a firm charging the highest price R are those that receive no other price quotes, $\bar{q}_1$ pins down the profits of all firms and hence the aggregate markup.

[39] Whether I assume free entry or an exogenous mass of firms has no effect on the level of markups in this model. As shown in Appendix C, markups are pinned down by consumer search behavior, and the free entry condition is cleared by changes in the mass of firms M. This would also be the case in a monopolistic competition model with CES preferences where the elasticity of substitution changes: a reduction in the elasticity of substitution results in higher equilibrium markups and thus an increase in the number of firms to maintain the zero profit condition.

It will also be useful to define a *stable equilibrium* in the model. Aggregating households' first-order conditions in (5) yields:[40]

$$\underbrace{\sum_{n=1}^{\infty}\left(\int_0^{\infty}\frac{-dQ_n(z)}{ds(z)}d\Lambda(z)\right)\left[\mathbb{E}\left[p|n\right]-\mathbb{E}\left[p|n+1\right]\right]}_{\text{Aggregate returns to search}} = \underbrace{\int_0^{\infty}\phi(z)d\Lambda(z)}_{\text{Aggregate cost of time}}\ , \tag{10}$$

where $\mathbb{E}[p|n]$ is the expected price paid after receiving n price quotes from F.

Consider how the aggregate returns to search change with $\bar{q}_1$. When $\bar{q}_1 = 1$, all households search only once, and hence firms identically set the monopoly price, $p = R$. On the other hand, when $\bar{q} = 0$, all households retrieve at least two price quotes, no firm can be incentivized to set a price above all other firms, and thus all firms price at marginal cost. In both cases, the price distribution is degenerate, and the returns to search are zero. For any intermediate value of $\bar{q}_1 \in (0, 1)$, however, the aggregate returns to search are strictly positive by Assumption 1(B).

I refer to an equilibrium in which aggregate returns to search are weakly increasing in $\bar{q}_1$ as a stable equilibrium and conduct all comparative statics locally around such an equilibrium. The intuition for stability is as follows. If aggregate returns to search are increasing in $\bar{q}_1$, an idiosyncratic increase in search effort exerted by any household leads to a decrease in $\bar{q}_1$, which decreases aggregate returns to search and hence leads to counteracting decreases in search intensity by other households. (In other words, household search decisions in the stable equilibrium are strategic substitutes.) In contrast, if aggregate returns to search are decreasing in $\bar{q}_1$, a perturbation in search effort by any household kicks off changes in search effort by all other households that lead away from the equilibrium point.

Since $Q_n(s)$ are continuous and smooth functions, if there exists a value of $\bar{q}_1$ such that the equilibrium condition (10) holds, then there must be at least one value of $\bar{q}_1$ where the left-hand side of (10) is weakly increasing in $\bar{q}_1$. Hence, if a dispersed-price equilibrium exists, there exists a stable dispersed-price equilibrium.

4.4 Search and markups paid in the cross-section

To characterize how search intensity and markups paid vary with income, I define two additional conditions on the mapping $\mathcal{S}$ from search intensity to the distribution of price quotes received.

Assumption 2. For any non-degenerate distribution F, the mapping $\mathcal{S} : s \mapsto \{q_n\}_{n=1}^{\infty}$ satisfies

$$\sum_{n=1}^{\infty}\frac{d^2Q_n}{ds^2}\left[\mathbb{E}\left[p|n;F\right]-\mathbb{E}\left[p|n+1;F\right]\right] > 0,$$

[40] I use the assumption here that all households have an internal solution for s_i. If the aggregate opportunity cost of search effort is too high, no dispersed-price equilibrium exists, and the sole equilibrium is the monopoly price equilibrium in which all households choose $s_i = 0$ and firms choose $p = R$.

where Q_n is the cumulative mass function of $\{q_n\}_{n=1}^{\infty}$ and where $\mathbb{E}[p|n;F]$ is the expected value of the minimum of n independent draws from the distribution F.

Assumption 3. For any non-degenerate distribution F, the mapping $\mathcal{S}$ satisfies

$$\sum_{n=1}^{\infty}\left(\frac{d^2Q_1}{ds^2}\frac{d^2Q_n}{ds^2}-\frac{dQ_1}{ds}\frac{d^3Q_n}{ds^3}\right)\left[\mathbb{E}\left[p|n;F\right]-\mathbb{E}\left[p|n+1;F\right]\right]\geq 0,$$

where Q_n and $\mathbb{E}[p|n;F]$ are as defined in Assumption 2.

While these two conditions may appear technical, their content is straightforward. Assumption 2 guarantees that returns to search diminish as search intensity increases. Assumption 3 is a sufficient condition such that, if $\phi(z)$ is increasing and convex in z, then the probability of receiving only one price quote $q_1(z)$ is also increasing and convex in z. I show in Appendix C.5 that both assumptions are satisfied by the two most common parameterizations of the Burdett and Judd (1983) model: (1) a version in which households receive only one or two quotes (e.g., Alessandria and Kaboski 2011, Pytka 2018, and Nord 2022), and (2) a version in which the number of price quotes received by households is drawn from a Poisson distribution (e.g., Albrecht et al. 2021 and Menzio 2021).

With these assumptions in hand, Lemma 2 shows how search intensity and markups paid vary in the cross-section of households.

Lemma 2 (Search intensity and markups paid)**.** *Suppose Assumption 2 holds. If the opportunity cost of search effort $\phi(z)$ is increasing (decreasing) in z, then*

1. *Search intensity $s(z)$ is decreasing (increasing) in z,*
2. *The average price $p(z)$ and markup paid $\mu(z)$ are increasing (decreasing) in z.*

Lemma 2 shows that markups paid may increase or decrease with labor productivity z, and that whether markups paid increase with z depends on whether search productivity $a(z)$ increases more or less than one-for-one with labor productivity. This means that, in principle, the model can generate either of two possibilities frequently posited in the literature: (1) if search productivity rises faster than one-for-one with labor productivity, low-income households receive fewer price quotes than high-income households in equilibrium and pay a "poverty premium" on average; (2) if search productivity rises less than one-for-one with labor productivity, high-income households exert less search intensity and hence pay higher average markups. Given the empirical evidence in Section 3.1, the latter case is the relevant one for our analysis.

4.5 Comparative statics of aggregate markup to $\Lambda(z)$

Proposition 1 provides sufficient conditions for a first-order stochastic shift in the distribution of buyers' incomes to increase the aggregate markup. Proposition 2 does the same for a mean-preserving spread (increase in inequality).

Proposition 1 (First-Order Shift in the Distribution of Buyers' Incomes)**.** *Suppose* $\Lambda(z)$ *is the distribution of buyers' incomes in a stable equilibrium. Consider a perturbation in the household distribution such that the resulting distribution* $\tilde{\Lambda}(z)$ *first-order stochastically dominates* $\Lambda(z)$*. Under Assumption 2, the aggregate markup in the economy with* $\tilde{\Lambda}(i)$ *is greater than the aggregate markup under* $\Lambda(z)$ *if* $\phi(z)$ *is increasing in* z*.*

Proposition 2 (Mean-Preserving Spread in the Distribution of Buyers' Incomes)**.** *Suppose* $\Lambda(z)$ *is the distribution of buyers' incomes in a stable equilibrium. Consider a perturbation in the household distribution such that the resulting distribution* $\tilde{\Lambda}(z)$ *is a mean-preserving spread of* $\Lambda(z)$*. Under Assumptions 2 and 3, the aggregate markup in the economy with* $\tilde{\Lambda}(i)$ *is greater than the aggregate markup under* $\Lambda(z)$ *if* $\phi(z)$ *is increasing and convex in* z*.*

Propositions 1 and 2 follow from Lemma 1. Since $\bar{q}_1$ is a sufficient statistic for the aggregate markup, the direct response of a change in the distribution of buyers' incomes on the aggregate markup depends on whether $q_1(z)$ is increasing (for Proposition 1) and convex (for Proposition 2) in z. The stability of the equilibrium then ensures that all indirect responses (i.e., the adjustment each household modifies its search intensity) do not overwhelm the direct response.

4.6 Discussion

Balanced growth. While Proposition 1 shows that a first-order stochastic shift in buyers' incomes leads to an increase in the aggregate markup, the model can generate a balanced growth path if increases in labor productivity are offset by increases in search productivity. For example, if labor productivity and search productivity for all households grow at the same rate, the distribution of ϕ is unchanged, and hence there is no change in the aggregate markup.

However, if labor productivity growth outpaces search productivity growth, Proposition 1 implies that the aggregate markup rises over time (with the economy tending in the limit toward the monopoly price equilibrium). This result relates to a literature that seeks to understand why markups and price dispersion have not fallen over time, despite apparent improvements in search technology. For example, Menzio (2021) shows that a decline in search frictions may coincide with constant price dispersion due to endogenous specialization by sellers. This paper makes the complementary point that increases in search productivity may not lead to a decline in markups and price dispersion if labor productivity also grows. In fact, due to the race between labor and search productivity, search productivity growth can coincide with *increasing* average markups.

Alternative models. Online Appendix E develops a version of the model that builds upon the model of sequential search by Burdett and Mortensen (1998), instead of the nonsequential search technology used here. That model delivers similar results qualitatively and quantitatively.

An alternative is to depart from the search micro-foundation altogether. Online Appendix F instead posits differences in elasticities of substitutes and taste for quality across households, following with Handbury (2021) and Faber and Fally (2022). Since elasticities of substitution and taste for quality are primitives of utility, that model does not feature a negative feedback loop between aggregate price sensitivity and each household's individual price sensitivity as in the model presented here.

4.7 Extension to heterogeneous products

The model as presented so far assumes that all firms produce perfectly substitutable varieties of a single consumption good. In taking the model to the data, I also consider an extension of the model that allows for heterogeneous goods. An advantage of this extension is that it can account for the wedge between the markup gap across income groups and the gap in prices paid for identical products across income groups. The primary disadvantage is that analytical results analogous to Propositions 1 and 2 are challenging to prove in the multiple-product case without strong restrictions on households' Engel curves across products.

To introduce heterogeneous products, I allow for a finite number of product segments K. Households' tastes for each product segment consist of an idiosyncratic component and a component that is correlated with household income. This micro-foundation relates to the model of niche consumption by Neiman and Vavra (2019), in which households draw potentially correlated taste shocks for different varieties. Formally, for each segment k, let v_{ik} denote household i's taste for segment k, given by

$$v_{ik} = \underbrace{\alpha_k(z_i)\delta_k}_{\substack{\text{Common to households} \\ \text{with income } z_i}} + \underbrace{\epsilon_{ik}}_{\text{Idiosyncratic}},$$

where ϵ_{ik} are independent draws from a Gumbel distribution. Given a household's tastes, I assume that household i receives utility only from the segment k for which it has the highest taste v_{ik}, and thus purchases exclusively from that segment.[41]

Aggregating across households, the fraction of households with income z that purchase from segment k is

$$\text{Share}_k(z) = \frac{\exp(\alpha_k(z)\delta_k)}{\sum_{k'} \exp(\alpha_{k'}(z)\delta_{k'})}.$$

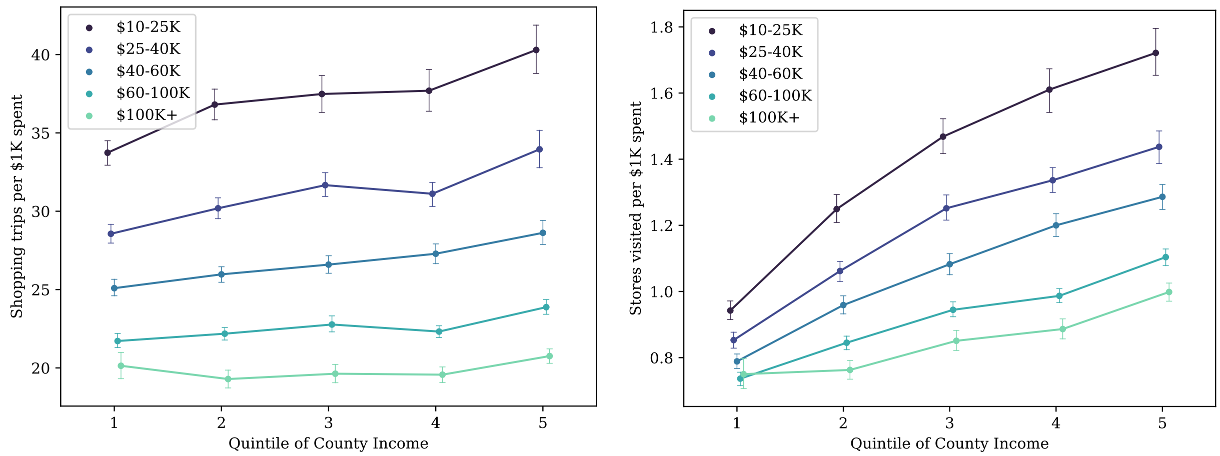
Note that by choosing $\alpha_k(z)$ and δ_k, we can match any observed shares of purchases made by households of income z in segment k.

The rest of the model is analogous to the single-good model presented above: each household chooses its search intensity to maximize utility, prices set by firms in each product segment follow a dispersed-price equilibrium, and free entry dictates the number of firms in each product segment. I relegate these details to Appendix C.6.[42]

[41] Formally, utility takes the form $u_i(\{c_{ik}\}) = \sum_k \mathbf{1}\{v_{ik} \geq v_{ik'},\ \forall k'\}\nu(c_{ik})$, where $\nu(\cdot)$ is strictly increasing. Assuming that a household only purchases from a single product segment is reasonable because, as pointed out by Handbury (2021), conditional on purchasing at least one product from a product module in a given month, households typically only purchase one product.

[42] This version of the model also accommodates more degrees of freedom—for example, marginal costs, entry costs, and households' reservation prices can vary flexibly by product segment—that may be desirable in matching features of the data. For the most part, these degrees of freedom are not essential to the counterfactuals I consider, and I silence

Figure 8: Shopping intensity by income group and county income.



(a) Shopping trips per $1K expenditures. **(b)** Unique stores visited per $1K expenditures.

5 Suggestive Evidence on Search Behavior

This section provides suggestive evidence for the search mechanism in the model. In particular, I show that measures of search intensity in the Nielsen data are consistent with two predictions of the model: (1) search intensity is decreasing in household income, and (2) conditional on income, search intensity is increasing in high-income areas.

I use two measures of shopping behavior previously used by Kaplan and Menzio (2015): the number of shopping trips a household makes and the number of unique stores it visits. I normalize both measures of shopping behavior by dollar spent to incorporate the insight by Pytka (2018) that search time reflects both search intensity and the size of the consumption basket; my objective is to isolate the former. Nevertheless, using expenditures to control for the size of the consumption basket risks confounding our results with differences in prices paid by income. Appendix Table B9 replicates the analysis instead normalizing by households' total number of transactions, number of unique UPCs purchased, or number of unique brands purchased. All three alternatives yield similar results to the ones presented here.

Figure 8 plots the two measures of shopping intensity—shopping trips per $1,000 expenditures (left panel) and unique stores visited per $1,000 expenditures (right panel)—across five income groups. The x-axis splits these income groups by the quintile of the average income in the county in which the household is based. Two patterns emerge. First, across all county quintiles, high-income households exert less search intensity per dollar spent. This fact has been previously established by Pytka (2018). Second, conditional on income, households exert greater search intensity in high-income counties. This finding is novel and is consistent with the model prediction that households search more when the search intensity of surrounding households decreases.

Table 8 formally tests this relationship. Columns 1 and 4 report that shopping intensity is

these additional degrees of freedom for ease of comparison across models.

decreasing in income: a 10 percent increase in household income is associated with 3–4 percent fewer shopping trips and stores visited per $1,000 spent. Columns 2 and 5 add average county income and use state (rather than county) fixed effects. Conditional on own income, shopping intensity increases with average county income. To account for the fact that a household's search productivity may vary across counties (e.g., due to store density or the quality of transportation available), columns 3 and 6 control for the number of grocery establishments in each county. While the coefficient on county income attenuates, it remains positive and significant. This finding is consistent with the model prediction that, conditional on income, a household will search more when surrounded by higher-income households.[43]

Table 8: Effect of income and county income on shopping intensity.

	Log Shopping Trips per $1K			*Log Unique Stores per $1K*		
	(1)	(2)	(3)	(4)	(5)	(6)
Log Household Income	−0.40**	−0.40**	−0.40**	−0.35**	−0.34**	−0.34**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log Avg. County Income		0.19**	0.10**		0.65**	0.27**
		(0.02)	(0.02)		(0.07)	(0.07)
Log Grocery Estabs.			0.03**			0.11**
			(0.00)			(0.01)
State FEs		Yes	Yes		Yes	Yes
County FEs	Yes			Yes		
N	63 350	62 865	62 859	63 350	62 865	62 859
R^2	0.14	0.09	0.09	0.23	0.11	0.14

Note: Household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $13.5K), and is instrumented using the education and occupation of the male and female heads of household. Average county income is from the BEA Personal Income by County Area release. Grocery Estabs. are a count of NAICS 445 establishments from Census Business Patterns (includes grocery stores, supermarkets, liquor stores, and specialty food stores). BEA and Census Business Pattern data is not available for a small number of counties in the Homescan sample. Standard errors clustered by county. * indicates significance at 10%, ** at 5%.

6 Calibration

In this section, I calibrate the model developed in Section 4. Calibrated parameters suggest significant spillovers across households from search behavior.

[43]Relatedly, Nevo and Wong (2019) find that returns to shopping declined during the Great Recession, even as households increased their search efforts; this is consistent with the model prediction that increases in household shopping intensity decrease all households' returns to search.

6.1 Calibration procedure

I assume that the mapping function $\mathcal{S} : s_i \mapsto \{q_{i,n}\}_{n=1}^{\infty}$ is Poisson as in Albrecht et al. (2021) and Menzio (2021), so that

$$q_{i,n+1} = \frac{s_i^n \exp(-s_i)}{n!}.$$

Saez and Zucman (2019) report income by percentile of the income distribution from 1950–2018, which are plotted in Appendix Figure B11. I take the equivalent of hourly wages z in the model to be proportional to average post-tax income. (I use post-tax rather than pre-tax income to account for the fact that a higher tax rate reduces the earnings from additional time spent working, and hence post-tax wages are the relevant proxy for households' opportunity cost of time.) For the baseline calibration, I set $dH(z)$ equal to the density of households by income in 2007 reported by Saez and Zucman (2019).

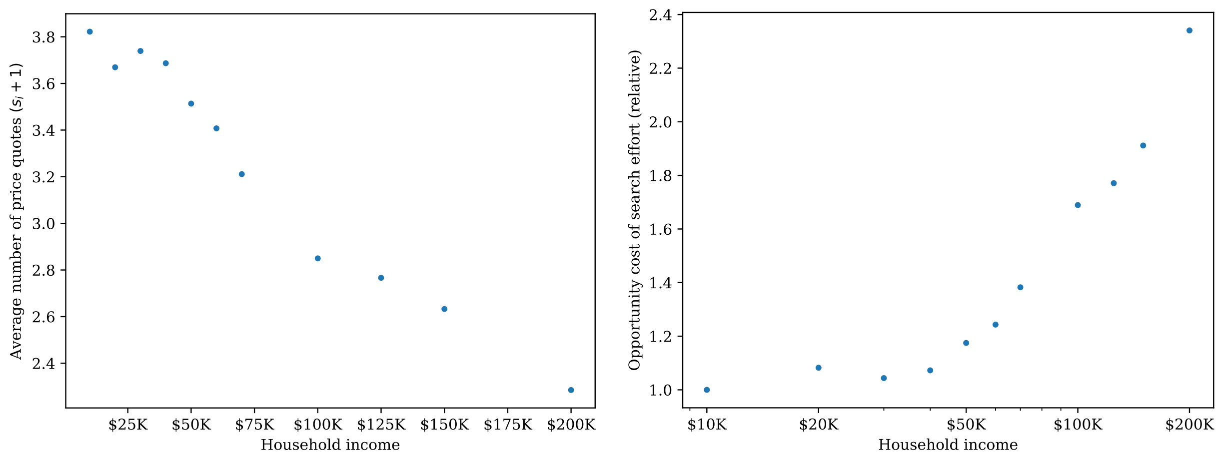
I normalize firms' marginal costs to one and set the reservation price to the 98th percentile of markups in the data, $R = 3.3$.[44] To calibrate the model with segmentation, I order all UPCs from lowest average buyer income to highest average buyer income and split the UPCs into K segments with equal sales. Appendix Figure B12 shows the fraction of expenditures by households at each income level on UPCs in each of the K segments for $K = 5$ and $K = 10$. As expected, low-income households disproportionately buy UPCs in the lowest segments, while high-income households disproportionately buy from the highest segments. I directly use the observed share of purchases made by each income group in each segment to calculate $\text{Share}_k(z)$.

I calibrate the remaining parameters on households' opportunity costs of search effort $\phi(z)$ and search productivities $a(z)$ to match the average markup paid by each income group (reported in Figure 1). I assume markups paid and search behavior of any household with income over \$200K are equal to the \$200K income group, so that the results are not influenced by extrapolation beyond the income range observed in the data. The calibration proceeds in two steps.

Inner loop: Offer distributions F_k and search intensities $s_k(z)$. Within each product segment, the distribution of posted prices F_k and the search behavior of all households shopping in that segment, $s_k(z)$, solve a fixed point. Given an initial guess of search intensities of each household in the segment, $s_k^{(0)}(z)$, the distribution of prices chosen by firms in that segment, $F_k^{(0)}$, is given by (7). The distribution $F_k^{(0)}$ then determines the returns to search at any level of search intensity. Hence, I use $F_k^{(0)}$ and the opportunity cost of search effort $\phi(z)$ to re-compute search intensity by income, $s_k^{(1)}(z)$. These search intensities yield a new price distribution $F_k^{(1)}$, and so on. I iterate the process until the price distribution $F_k^{(t)}$ and search intensities $s_k^{(t)}(z)$ each converge. By running this procedure with different initial conditions, I confirm that the within-segment equilibrium price distribution is unique.

[44] Appendix Table B13 reports how the main results change with R. A decrease in R moderately increases the model's prediction of changes in markups over time, so setting $R = 3.3$ is conservative.

Figure 9: Price quotes received and opportunity cost of search effort, for $K = 1$.



(a) Average no. of price quotes received ($s(z) + 1$). **(b)** Oppty. cost $\phi(z)$ (normalized to one for <$20K).

Outer loop: Opportunity costs of search effort $\phi(z)$ and search productivities $a(z)$. The inner loop is nested in an outer loop, which calculates the opportunity cost of search effort for each income group. Given an initial guess $\phi^{(0)}(z)$, I calculate the price distributions F_k and search intensities $s_k(z)$ in all segments. For each income level z, given shares of purchases in each segment $\text{Share}_k(z)$ and the price distributions F_k, there is a one-to-one mapping between opportunity cost ϕ and the average markup paid across segments, $\mu(z)$. Thus, I choose $\phi^{(1)}(z)$ to match the average markup paid by each income group in Figure 1. New price distributions F_k and search intensities $s_k(z)$ are then re-computed using the new $\phi^{(1)}(z)$. I iterate this process until the calibrated values $\phi^{(t)}(z)$ converge. Finally, I calculate search productivities using $a(z) = z/\phi(z)$.

6.2 Calibrated statistics

Figure 9 plots calibrated statistics on search intensity and opportunity cost of search effort by income in the single-product model. (Appendix Figure B13 reports analogous results for the model with $K = 10$ segments.) The left panel shows that the expected number of price quotes, $(s(z) + 1)$, decreases with income: households with over $200K in income receive on average 40 percent fewer price quotes per purchase than low-income households. The calibrated opportunity cost of search effort $\phi(z)$, shown in the right panel, is increasing and convex in log income, satisfying the conditions in Propositions 1 and 2.[45]

[45] In the calibration with $K = 10$ segments, the calibrated statistics are similar: high-income households receive 30 percent fewer price quotes and have a opportunity cost of search effort 80 percent higher than low-income households.

Untargeted moments. Table 9 compares the elasticity of markups paid and search intensity to own and others' incomes in the model to the data.[46] Since the model is calibrated to match the average markup paid by income group, the model matches the elasticity of markups paid to own income in the data. The elasticity of markups paid to others' incomes is somewhat conservative relative to the data. As a result, the macro elasticity of markups to income for the single-product model (9.0%) is on the low end of the range estimated in the data. Spillovers are limited further in the model with segmentation, and the macro elasticity of markups to income in the model with $K = 10$ segments (6.8%) undershoots the target range.

The elasticity of the average number of price quotes received to income in the model ranges from −0.11 to −0.17, within the range of −0.08 to −0.40 estimated in the data (see Appendix Table B9). The extent of the difference in search intensity between high- and low-income households is also consistent with previous work by McKenzie and Schargrodsky (2005), who measure shopping frequency for households in Argentina and find that shopping frequency for households at the ninetieth percentile of household income is 30 percent lower than that of households at the tenth decile, controlling for quantity purchased. Their estimates correspond to an elasticity of about −0.12. Additionally, the model generates a positive elasticity of search intensity to others' incomes, consistent with the findings in Section 5.

As an additional check, I compare the elasticity of the price search function $p(s; F)$ to previous work. In the model, doubling time spent shopping decreases prices paid by 7.5–9.2%.[47] This estimate is in line with Aguiar and Hurst (2007a), who estimate that doubling shopping frequency lowers prices paid by 7–10 percent.

In the model with heterogeneous products, differences in buyer composition across segments create a wedge between the difference in prices paid within a product segment and the difference in markups paid unconditionally across income groups. Quantitatively, in the calibration with $K = 10$ segments, high-income households pay 7.5 percent higher prices than low-income households within each product segment. This is in line with the magnitude of the gap in prices paid for identical products in the data (6–7.5 percent). (Appendix Figure B15 plots the within-segment gap in prices paid against estimates from the data for each income group.) Hence, the model is able to simultaneously match differences in markups paid across households with a smaller gap in prices paid for identical products.

Finally, in Appendix Table B10, I compare various percentiles of the markup distribution in the data to those predicted by the single-product model for four income groups. While there are some differences of note (for example, in the data, more than 10 percent of purchases made by households are made at markups below one, while the model does not admit prices below marginal cost), the model appears to fit the data reasonably well.

[46] To calculate the elasticity of markups paid and search behavior to others' income, I calibrate the model using the income distribution from various years from 1950–2018, as in Section 7. This allows me to estimate regressions of markup paid with respect to own income as well as the average per-capita income across these economies.

[47] The range corresponds to differences across low- and high-income households. Appendix Figure B14 plots the expected price paid against search intensity for the single-product model.

Table 9: Elasticity of markups paid and search intensity to own and others' incomes in the data and in the model.

	Log markup				*Search behavior*			
	Data		Model		Data		Model	
	OLS	IV	$K = 1$	$K = 10$	OLS	IV	$K = 1$	$K = 10$
Log Household Income	0.034	0.053	0.034	0.032	−0.26	−0.40	−0.17	−0.11
Log Others' Income	0.101	0.092	0.056	0.036	0.03	0.10	0.06	0.03

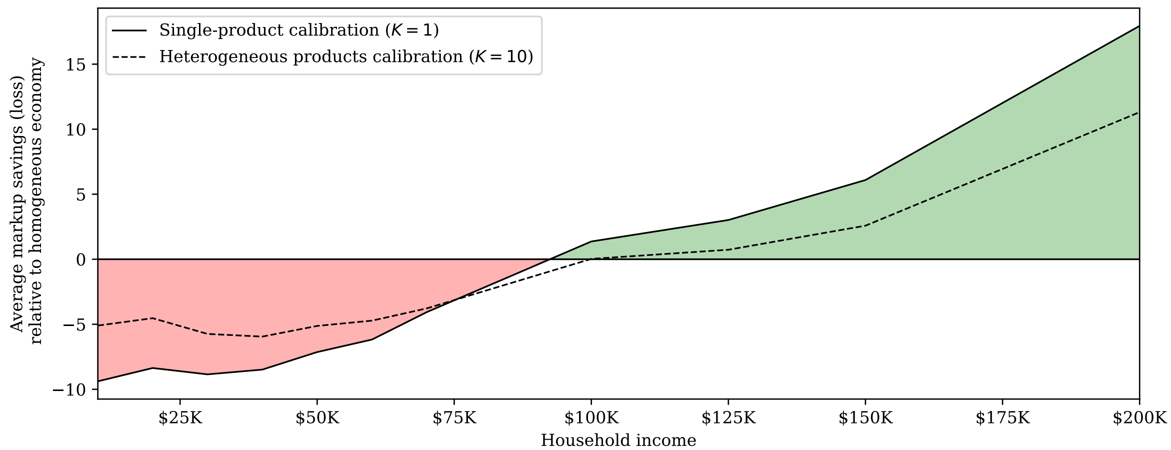
Note: The first and second column are from Table 4 columns 2 and 3. The third and fourth columns report results from a regression of log markup paid on log household income and log average income in the calibrated model. The fifth and sixth columns report OLS and IV estimates of the elasticity of shopping trips per dollar spent to household income and county income, as in Table 8. The seventh and eighth column report results from a regression of avg. price quotes received per purchase on log household income and log average income in the model.

Comparison to previous estimates of price sensitivity. Prior work estimating differences in price sensitivity across households (such as Faber and Fally 2022, Handbury 2021, Gupta 2020, and Auer et al. 2022) often estimate elasticities of substitution. To compare my estimates to theirs, I calculate the aggregate markup that would result in an economy with homogeneous households of each income level and compute the elasticity of substitution $\sigma(z) = \frac{\mu^{\text{homog}}(z)}{\mu^{\text{homog}}(z)-1}$ that generates this markup. In the single-product model, these elasticities of substitution decrease from 6 for the lowest-income households to 2.5 for the highest-income households, and are plotted against estimates from Auer et al. (2022) in Appendix Figure B16. The differences in price sensitivity I estimate are larger than those estimated by Faber and Fally (2022), who find that price elasticities of the two lowest-income quintiles are only 0.4 higher than that of the richest quintile, but are of a similar magnitude as those estimated by Handbury (2021) and Auer et al. (2022).

Spillovers. In the model, households' search decisions affect the distribution of prices offered by firms and hence affects the markups paid by other households. To quantify these spillovers, for each income level I consider an economy in which there is a representative household with labor productivity z and search productivity $a(z)$. I compare the average markup that each household would pay in that economy to the average markup it pays in the baseline calibration.

Figure 10 shows that the difference in average markups paid are large in magnitude: low-income households would pay 5–9pp lower markups in an economy populated with only low-income households. On the other hand, households with over $200K in income save over 10pp on markups paid due to the greater search intensity of low-income households. These differences in markups paid are also accompanied by differences in search behavior: low-income households retrieve as many as 10 percent more price quotes than they would in an economy composed of only low-income households, while high-income households retrieve 2 percent fewer price quotes when surrounded by low-income households. These spillovers are responsible for the wedge between the micro and macro elasticities of markups to income recorded in Table 9.

Figure 10: Savings (losses) in average markup paid relative to homog. income economy.



7 Changes in the Income Distribution from 1950–2018

How do changes in the income distribution over time affect the aggregate retail markup? In this section, I first provide back-of-the-envelope estimates using moments from the data, which suggest that income growth from 1950–2018 predicts an increase in the aggregate retail markup between 5 and 23pp. I then use the model calibrated in the previous section. Through the lens of the model, changes in the income distribution from 1950–2018 account for a 10–14pp rise in the aggregate retail markup.

7.1 Back-of-the-envelope estimates

I start by using moments directly from the data to do a back-of-the-envelope calculation on how changes in the income distribution over time affect the aggregate retail markup.

First, consider the case of perfect price discrimination, in which case each household pays a certain markup irrespective of the incomes of other households. Compositional changes in the income distribution from 1950–2018 then result in a 5.7pp increase in the average retail markup from 1950–2018.[48]

Section 3.3, however, presents evidence against perfect price discrimination. Accordingly, empirical evidence in that section places the macro elasticity of markups to income between 8–15%. Since post-tax per-capita real income grew 3.5 times from 1950 to 2018, with an average markup of 32 percent, a back-of-the-envelope calculation suggests a change in the aggregate markup between 13–23pp:

$$1.32 \times \log(3.5) \times (0.08\text{-}0.15) = 13.2\text{–}23.2\text{pp}$$

[48]Similarly, for the case of perfect price discrimination, we could assume that the macro elasticity of markups is exactly equal to the micro elasticity observed in the cross-section. Using a micro elasticity of 3.1%, this predicts an increase in the aggregate markup of $1.32 \times \log(3.5) \times 0.031 = 5.1$pp.

These estimates do not account for rising income dispersion, which, following Proposition 2, would predict a greater change in the retail markup over time. However, the empirical estimates of macro elasticities of markup to income may not be able to fully capture strategic interactions in search intensity or market segmentation, which would tend to decrease the predicted change in the retail markup over time. Hence, I now turn to the model.

7.2 Model estimates

The relevant statistic for how changes in the income distribution affect the aggregate retail markup is the macro elasticity of markups to income. Since the single-product model most closely matches this moment in the data, I use the single-product model as the baseline for this analysis. I discuss below how results differ when using the model with heterogeneous products.

To calculate the aggregate retail markup predicted by the model under income distributions from 1950–2018, I assume that households with real post-tax earnings z have the same search productivity $a(z)$ and hence opportunity cost of search effort $\phi(z)$ as households with the same real post-tax earnings in 2007.[49] I then solve for the distribution of offered prices F and the aggregate retail markup $\bar{\mu}$ under the new income distribution.

The solid blue line in Figure 11 plots the predicted aggregate retail markup over time using the income distributions from 1950–2018 from Saez and Zucman (2019). Over this period, the model predicts a 14pp increase in the aggregate retail markup. The rise in markups is mild from 1950–1980 but accelerates significantly from 1980–2000.

To understand the degree to which the rise in retail markups is due to changes in the dispersion, rather than the level, of income, the dotted black line in Figure 11 plots the predicted retail markups holding income dispersion constant at 1950 levels. The change in predicted markups before 1980 is nearly identical to the change predicted under the realized income distribution. However, the two series diverge in 1980 as income dispersion rises. In 2018, the predicted markup at the 1950 level of income dispersion is 3.7pp lower than at the 2018 level of income dispersion. Table 10 summarizes the predicted change in markups and the portion due to rising income dispersion.

The role of reallocations. Changes in the aggregate markup may reflect both changes in the markups set by individual firms and a compositional shift reallocating sales toward high-markup firms. Autor et al. (2020) and Kehrig and Vincent (2021) suggest that reallocation across firms has played the dominant role in increasing markups (and decreasing the labor share) in the U.S. economy. On the other hand, Döpper et al. (2021) find that an increase in markups estimated in Nielsen scanner data is driven primarily by changes within products over time.

[49] Of course, it is possible that search productivity has changed secularly over time (e.g., due to the introduction of e-commerce). However, this would imply that the time series elasticity of markups to income should differ significantly from the cross-sectional elasticity, a prediction not borne out in Section 3.3. In Online Appendix D, I explore how the search productivity growth baked into the model compares to the empirical evidence. The results suggest that changes in search behavior over time in the model align well with the data.

Figure 11: Predicted aggregate retail markup under income distributions from 1950–2018.

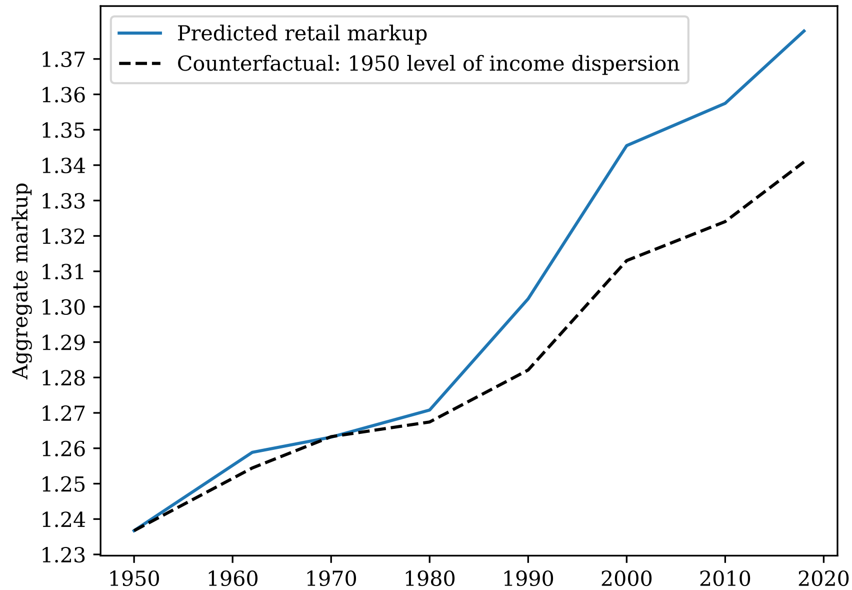


Table 10: Predicted change in aggregate retail markup from 1950–2018.

Period	Predicted Δ in markup	Due to Δ Income level	Due to Δ Income dispersion	Due to Within-product changes	Due to Cross-product reallocations
1950–2018	14.2pp	10.5pp	3.7pp	10.1pp	4.2pp
1950–1980	3.4pp	3.1pp	0.3pp	2.5pp	1.0pp
1980–2018	10.8pp	7.4pp	3.3pp	7.7pp	3.1pp

Table 10 reports that about 70 percent of the increase in the aggregate retail markup over time in the model comes from changes in markups for individual products, while 30 percent is due to reallocations across products. To see this breakdown over time, Appendix Figure B18 decomposes the change in the aggregate markup from 1950–2018 into these two channels.[50] The reallocation to products at higher quantiles of the markup distribution occurs because high-income households search less and are therefore more likely to buy from firms with relatively high markups. As a result, from 1950 to 2018, the model predicts that sales of firms at the bottom of the markup distribution fall 15 percent, while the sales of firms at the top of the markup distribution grow over 30 percent.

[50]Formally, the aggregate markup in period t is the sales-weighted harmonic average of all firm markups, $\bar{\mu}_t = \mathbb{E}[\lambda_{i,t}/\mu_{i,t}]^{-1}$, where $\lambda_{i,t}$ is the sales share of firm i at time t. The change in the aggregate markup from $\bar{\mu}_0$ to $\bar{\mu}_1$ is decomposed as the change in within-firm markups ($\mathbb{E}[\lambda_{i,0}/\mu_{i,1}]^{-1} - \mathbb{E}[\lambda_{i,0}/\mu_{i,0}]^{-1}$) and the change due to reallocations across firms ($\mathbb{E}[\lambda_{i,1}/\mu_{i,1}]^{-1} - \mathbb{E}[\lambda_{i,0}/\mu_{i,1}]^{-1}$).

Effect of strategic interactions in search behavior. Appendix Figure B17 shows that, as aggregate income rises, households each individually exert more search effort. These strategic interactions in search behavior moderate the change in the aggregate markup over time. If search behavior is instead held constant at 2007 levels, the model would predict a larger increase (19pp) in the aggregate markup this period (see Appendix Table B12).

The presence of these strategic interactions in search behavior also explains why the prediction of this model is more conservative than an alternative model in which differences in markups paid across households arise due to utility primitives, such as differences in taste for quality and elasticities of substitution (see the model developed in Online Appendix F). Since utility primitives are static over time, those models do not generate a negative feedback loop between aggregate income and price sensitivity, which dampens the overall change in markups over time.

How important is segmentation? While incorporating heterogeneous products allows the model to capture the wedge between differences in markups and differences in prices paid for identical products, it also shrinks the macro elasticity of markups to income below the range of estimates from the data. Accordingly, the change in the aggregate markup predicted over time by the model with heterogeneous product is also smaller. Appendix Table B11 reports the predicted change in the aggregate retail markup over time, varying the number of product segments from $K = 1$ to 100. The predicted change in the aggregate markup is between 10–11pp.[51,52]

How does the rise in markups compare to data? Figure 12 compares the path of the aggregate markup in the model to data on gross margins for grocery stores from the Census of Annual Retail Trade Statistics.[53] An aggregate markup calculated from grocery store gross margins reported by the Census increases from 29 percent in 1983 to 38 percent in 2020. The path of the aggregate retail markup predicted by the model appears to fit this upward trend in markups in the data well.

8 Conclusion

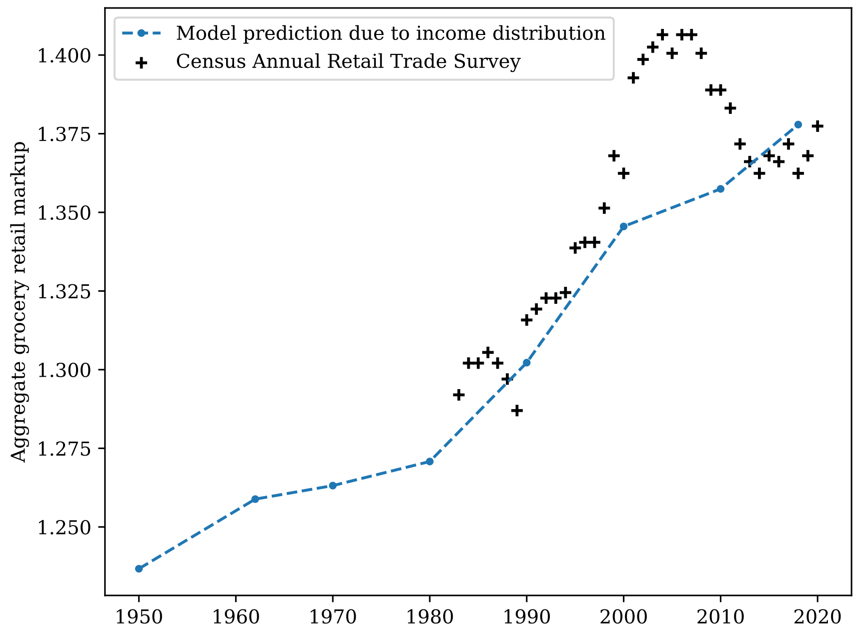
Price elasticities facing firms depend both on the presence of alternatives supplied by competing firms and the propensity of consumers to switch to these alternatives. This paper explores one

[51] Recent work by Nord (2022) compares the prices paid by households of different incomes to prices they would pay in counterfactuals where segmentation is entirely absent and where segmentation is perfect. In line with the results presented here, he finds that prices paid in the baseline more closely resemble the case where segmentation is absent rather than the case of perfect price discrimination.

[52] One may additionally wonder how allowing segmentation to change endogenously affects the results. I consider an extension in which households substitute across segments over time according to a non-zero elasticity with respect to expected price paid. Varying this elasticity of substitution across segments from zero to ten results in less than a 0.1pp change in the predicted change in markups from 1950 to 2018.

[53] Grocery stores are most analogous to the set of retailers captured in the Nielsen data. However, Appendix Figure B19 show that other retailer categories in the Census of Annual Retail Trade Statistics show a similar upward trend in gross margins over time. Historical data on retail gross margins from an NBER study by Barger (1955) also suggests that the upward trend in markups extends farther back in time; he reports an aggregate markup for retail grocery stores in the range of 21–24 percent from 1869–1947.

Figure 12: Comparison of markups predicted by model to data on retail grocery gross margins.



Note: Average gross margins (total sales less total cost of goods sold, as a percent of sales) for retail grocery stores are available from the Census Annual Retail Trade Survey from 1983 to 2020. Under constant returns, the aggregate markup is given by Agg. Markup = Sales/Costs = 1/(1 − Gross Margin).

salient consumer characteristic—income—that affects a consumer's propensity to search for and substitute across products. Empirical evidence from retail markups suggests that high-income households tend to be less price sensitive, as conjectured by Harrod (1936), and that firms selling to high-income households thus set higher markups. This evidence suggests that markups are not purely a supply-side phenomenon: rather, changes to the composition of demand can play an important role in affecting the level and markups charged by firms.

This paper's focus is on the time series of the U.S. aggregate markup. Differences in the level and distribution of income across geographies or over the business cycle may also affect markups across space and over time. I am pursuing these extensions in ongoing work.

References

Afrouzi, H., A. Drenik, and R. Kim (2021). Growing by the masses. Working paper.

Aguiar, M. and E. Hurst (2007a). Life-cycle prices and production. *American Economic Review 97*(5), 1533–1559.

Aguiar, M. and E. Hurst (2007b). Measuring trends in leisure: The allocation of time over five decades. *The Quarterly Journal of Economics 122*(3), 969–1006.

Aguiar, M., E. Hurst, and L. Karabarbounis (2013). Time use during the great recession. *American Economic Review 103*(5), 1664–96.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics 109*(3), 659–684.

Albrecht, J., G. Menzio, and S. Vroman (2021). Vertical differentiation in frictional product markets. ' 29618, National Bureau of Economic Research.

Alessandria, G. and J. P. Kaboski (2011). Pricing-to-market and the failure of absolute ppp. *American Economic Journal: Macroeconomics 3*(1), 91–127.

Anderson, E., S. Rebelo, and A. Wong (2018). Markups across space and time. Technical Report 24434, National Bureau of Economic Research.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Auer, R. A., A. Burstein, S. Lein, and J. Vogel (2022). Unequal expenditure switching: Evidence from switzerland. Technical Report 29757, National Bureau of Economic Research.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2017). Concentrating on the fall of the labor share. *American Economic Review 107*(5), 180–85.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Barger, H. (1955). *Distribution's Place in the American Economy Since 1869*, Chapter "Trends in Margins", pp. 80–90. NBER.

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance 75*(5), 2421–2463.

Baye, M. R., J. Morgan, and P. Scholten (2004). Price dispersion in the small and in the large: Evidence from an internet price comparison site. *The Journal of Industrial Economics 52*(4), 463–496.

Bertoletti, P. and F. Etro (2017). Monopolistic competition when income matters. *The Economic Journal 127*(603), 1217–1243.

Bewley, T. (2007). Report on an ongoing field study of pricing as it relates to menu costs. Technical report, Cowles Foundation Conference, "The Macroeconomic of Lumpy Adjustment".

Bhardwaj, A., D. Ghose, S. Mukherjee, and M. Singh (2022). Million dollar plants and retail prices. Policy Research Working Paper 9995, World Bank.

Brand, J. (2021). Differences in differentiation: Rising variety and markups in retail food stores. Working Paper.

Broda, C., E. Leibtag, and D. E. Weinstein (2009). The role of prices in measuring the poor's living standards. *Journal of Economic Perspectives 23*(2), 77–97.

Broda, C. and D. E. Weinstein (2010). Product creation and destruction: Evidence and price implications. *American Economic Review 100*(3), 691–723.

Brynjolfsson, E. and M. D. Smith (2000). Frictionless commerce? a comparison of internet and conventional retailers. *Management Science 46*(4), 563–585.

Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica 51*(4), 955–969.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.

Coibion, O., Y. Gorodnichenko, and G. H. Hong (2015). The cyclicality of sales, regular and effective prices: Business cycle and policy implications. *American Economic Review 105*(3), 993–1029.

De Loecker, J., J. Eeckhout, and S. Mongey (2021). Quantifying market power and business dynamism in the macroeconomy. Technical Report 28761, National Bureau of Economic Research.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

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

Diamond, P. (1971). A model of price adjustment. *Journal of Economic Theory 3*, 156–68.

Diamond, R. and E. Moretti (2021). Where is standard of living the highest? local prices and the geography of consumption. Technical Report 29533, National Bureau of Economic Research.

Dickson, P. R. and A. G. Sawyer (1990). The price knowledge and search of supermarket shoppers. *Journal of Marketing 54*(3), 42–53.

Döpper, H., A. MacKay, N. H. Miller, and J. Stiebale (2021). Rising markups and the role of consumer preferences. Technical report, Working Paper.

Faber, B. and T. Fally (2022). Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. *The Review of Economic Studies 89*(3), 1420–1459.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011). Income distribution, product quality, and international trade. *Journal of Political Economy 119*(4), 721–765.

Gopinath, G., P.-O. Gourinchas, C.-T. Hsieh, and N. Li (2011). International prices, costs, and markup differences. *American Economic Review 101*(6), 2450–86.

Griffith, R., E. Leibtag, A. Leicester, and A. Nevo (2009). Consumer shopping behavior: How much do consumers save? *Journal of Economic Perspectives 23*(2), 99–120.

Gupta, A. (2020). Demand for quality, variable markups and misallocation: Evidence from india. Working paper.

Gutiérrez, G. (2017). Investigating global labor and profit shares. Working Paper.

Gutiérrez, G. and T. Philippon (2018). How eu markets became more competitive than us markets: A study of institutional drift. Technical Report 24700, National Bureau of Economic Research.

Handbury, J. (2021). Are poor cities cheap for everyone? non-homotheticity and the cost of living across u.s. cities. *Econometrica 89*(6), 2679–2715.

Harrod, R. F. (1936). *Trade Cycle. An essay.* London: Oxford University Press.

Hoch, S. J., B.-D. Kim, A. L. Montomery, and P. E. Rossi (1995). Determinants of store-level price elasticity. *Journal of Marketing Research 32*(1), 17–29.

Hummels, D. and V. Lugovskyy (2009). International pricing in a generalized model of ideal variety. *Journal of Money, Credit and Banking 41*(1), 3–33.

Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the us retail sector. *The Quarterly Journal of Economics 134*(2), 715–783.

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

Kaplan, G. and G. Menzio (2016). Shopping externalities and self-fulfilling employment fluctuations. *Journal of Political Economy 124*(3), 771–825.

Kehrig, M. and N. Vincent (2021). The micro-level anatomy of the labor share decline. *The Quarterly Journal of Economics 136*(2), 1031–1087.

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

Lach, S. (2007). Immigration and prices. *Journal of Political Economy 115*(4), 548–587.

McKenzie, D. and E. Schargrodsky (2005). Buying less, but shopping more: Changes in consumption patterns during a crisis. Technical Report 92, Bureau for Research and Economic Analysis of Development.

Menzio, G. (2021). Optimal product design: Implications for competition and growth under declining search frictions. Technical Report 28638, National Bureau of Economic Research.

Nakamura, E. and D. Zerom (2010). Accounting for incomplete pass-through. *Review of Economic Studies 77*, 1192–1230.

Neiman, B. and J. Vavra (2019). The rise of niche consumption. Technical Report No. 26134, National Bureau of Economic Research.

Nevo, A. and A. Wong (2019). The elasticity of substitution between time and market goods: Evidence from the great recession. *International Economic Review 60*(1), 25–51.

Nord, L. (2022, August). Shopping, demand composition, and equilibrium prices. Working paper.

Pisano, L., L. Riva, and A. Stella (2022). Price heterogeneity and consumption inequality. Working paper.

Pratt, J. W., D. A. Wise, and R. Zeckhauser (1979). Price differences in almost competitive markets. *The Quarterly Journal of Economics 93*(2), 189–211.

Pytka, K. (2018). Shopping effort in self-insurance economies. Working Paper.

Saez, E. and G. Zucman (2019). *The triumph of injustice: How the rich dodge taxes and how to make them pay*. WW Norton and Company.

Scholten, P. and S. A. Smith (2002). *The Economics of the Internet and E-commerce*, Chapter Price dispersion then and now: Evidence from retail and e-tail markets. Emerald Group Publishing.

Simonovska, I. (2015). Income differences and prices of tradables: Insights from an online retailer. *The Review of Economic Studies 82*(4), 1612–1656.

Sommeiller, E., M. Price, and E. Wazeter (2016). Income inequality in the us by state, metropolitan area, and county. Technical report, Economic Policy Institute.

Stroebel, J. and J. Vavra (2019). House prices, local demand, and retail prices. *Journal of Political Economy 127*(3), 1391–1436.

Varian, H. R. (1980). A model of sales. *American Economic Review 70*(4), 651–659.

Online Appendix

(Not for publication)

Appendix A Data Cleaning and Construction

A.1 Nielsen Homescan

Following the Nielsen data manual, I exclude magnet products (fresh produce and other items without barcodes) from my analysis.

Continuous measure of household income. As noted by Broda et al. (2009), Nielsen reports income in discrete categories. For all analyses that include household income as a continuous variable, I follow Broda et al. (2009) and recode each household's income as the midpoint of the income bracket. For example, a household earning $13,000 is part of the $12,000-$15,000 income group and is assigned an income equal to $13,500. For the group with over $200,000 in annual income, I assign an income of $225,000.

Merge with GS1. I merge UPCs from the Nielsen data with data on company ownership from Global Standards One (GS1), which supplies firms with unique company and product barcodes. Over 98.5 percent of transactions are merged with a parent company through GS1. I calculate a firm's sales share in a product module as the sales of all UPCs owned by the parent company divided by total product module sales.

Relative unit prices. I define the relative unit price $\hat{p}_g$ for transaction g as the difference between the log unit price paid in transaction g and the log average unit price paid for all transactions in the same product module measured in the same units (such as ounces, pounds, or count):

$$\hat{p}_g = \log\left(\frac{\text{Price}(g)}{\text{Units}(g)}\right) - \log\left(\frac{\sum_{g' \in \text{Module}(g)} \text{Price}(g')}{\sum_{g' \in \text{Module}(g)} \text{Units}(g')}\right). \tag{11}$$

In the above expression, Module(g) is the subset of products in the same product module as g measured in the same units as g, Price(g) is the price paid in transaction g, and Units(g) are the total units purchased in transaction g (the quantity of items sold times the ounces/pounds/count/etc. per item). Products with a high price index $\hat{p}$ are those that sell at higher unit prices than the average product in their product module.

Treatment of retailer IDs. Nielsen provides a retailer code for each transaction in the data, which designate retail chains (the identities of the retailers are anonymized for privacy reasons). Some retailer IDs, however, are catch-all IDs, which means they capture all remaining retail chains not uniquely coded by Nielsen. For all analyses that use retailer fixed effects, I use the subsample of transactions for which the retailer IDs uniquely identifies a retail chain. For the robustness analysis removing the largest retailers from the sample in Section 3.1, I rank uniquely identified retail chains by total sales observed in the Nielsen Homescan data. All retail chains that are not uniquely identified in the Nielsen data are assumed to be smaller than those uniquely identified.

A.2 PromoData Price-Trak

Wholesale costs provided by PromoData include a list of active categories and inactive categories (which PromoData uses to update an internal product encyclopedia). Following the data manual, I use both the active and inactive databases and drop duplicated observations in the inactive database. The database includes about 114,000 UPCs that also show up in the Nielsen Homescan data. Each UPC may be listed multiple times since it is available in different pack sizes to retailers; I call each unique UPC-pack size available to a retailer an "item" in the following description.

Data cleaning and construction. I construct the monthly wholesale base price and deal price for each item-market pair as the minimum reported base and deal prices for the item in the market in each month. Of about 260,000 items, wholesale prices for about 59,000 items are observed in at least two markets in a given month. Let $w_{i,m,t}^{\text{base}}$ and $w_{i,m,t}^{\text{deal}}$ be the wholesale base price and deal price of item i in market m in month t. I calculate the relative price $\hat{w}_{i,m,t}^{x}$ as the ratio of the price of i in market m to the modal price for i across markets in t. Consistent with Stroebel and Vavra (2019), I find that wholesale prices are surprisingly uniform across markets: Table A1 shows that over 80 percent of items in a given month in 2009 have a wholesale cost exactly equal to the modal price across markets.[54]

I assume that retailers purchase UPCs at the minimum price available to them, and so I calculate wholesale base and deal prices for each UPC in each month by taking the minimum price at which the UPC is offered across items (pack sizes) in that month. Since the PromoData lack information on the quantities of each item sold, this is a more principled approach than taking an unweighted average across items.

Merge with Nielsen Homescan. I merge these monthly wholesale costs into the Homescan data using the date of the shopping trip recorded by the panelist and the product UPC. As a check, I calculate the sales-weighted average markup for each UPC over all purchases observed in the Homescan data. Table A2 shows summary statistics on the distribution of UPC markups. Over 90 percent of UPCs have markups that lie between one and 2.5, and the fraction of products with an average markup below one is under 10 percent.

Table A3 reports the match rate of wholesale cost data by income group. The percent of transactions matched to wholesale cost data increases slightly with income, and the share of expenditures matched to wholesale cost data decreases slightly with income.

To check whether transactions matched to wholesale cost data are similar to the unmatched transactions, I compare the average of the relative price indices (as defined in (11)) for matched and unmatched transactions by income group. The final two columns of Table A3 show the average price index for matched and unmatched products by income group. We see that for middle- and high-income groups, the average price index on unmatched products is similar to the average

[54] Stroebel and Vavra (2019) conduct a similar analysis at a quarterly level across all years using a subset of 32 markets in the wholesale cost data, and find a similar figure of 78 percent.

Table A1: Uniformity of wholesale prices across markets.

	Measure of wholesale cost	
	Base Price	Deal Price
Percent of items sold:		
At modal price ($\hat{w}^x_{i,m,t} = 1$)	80.3	78.5
Within 5% of modal price ($\lvert\hat{w}^x_{i,m,t} - 1\rvert \leq 0.05$)	90.7	86.4
Within 10% of modal price ($\lvert\hat{w}^x_{i,m,t} - 1\rvert \leq 0.10$)	95.1	90.9

price index on matched products. For the lowest income groups, however, unmatched products tend to have lower price indices than those matched to the wholesale cost data. To the degree that the price index of a product covaries positively with its markup, this means that differences in markups calculated for our matched sample will be conservative relative to the true differences in markups across income groups.

To check whether using uniformity of wholesale prices to extend the sample materially affects the results, I replicate my baseline analyses using PromoData wholesale prices matched only to transactions made by households living in the same market as the wholesaler. I use a hand-constructed crosswalk from Scantrack Market IDs in the Nielsen Homescan panel to PromoData Price-Trak market areas. Of 63,350 panelists in the 2007 Nielsen Homescan panel, 30,922 (49%) live in markets for which there is a corresponding PromoData Price-Trak market area (though some PromoData market areas also have relatively scarce coverage of UPCs). Overall, the sample matched at the wholesale market level includes 3.0 million transactions (1.8 million of which are at stores with unique Nielsen IDs). The analyses replicated using this subset of the data are reported in Figure B5 and Table B2.

Table A2: Summary statistics for markup distribution.

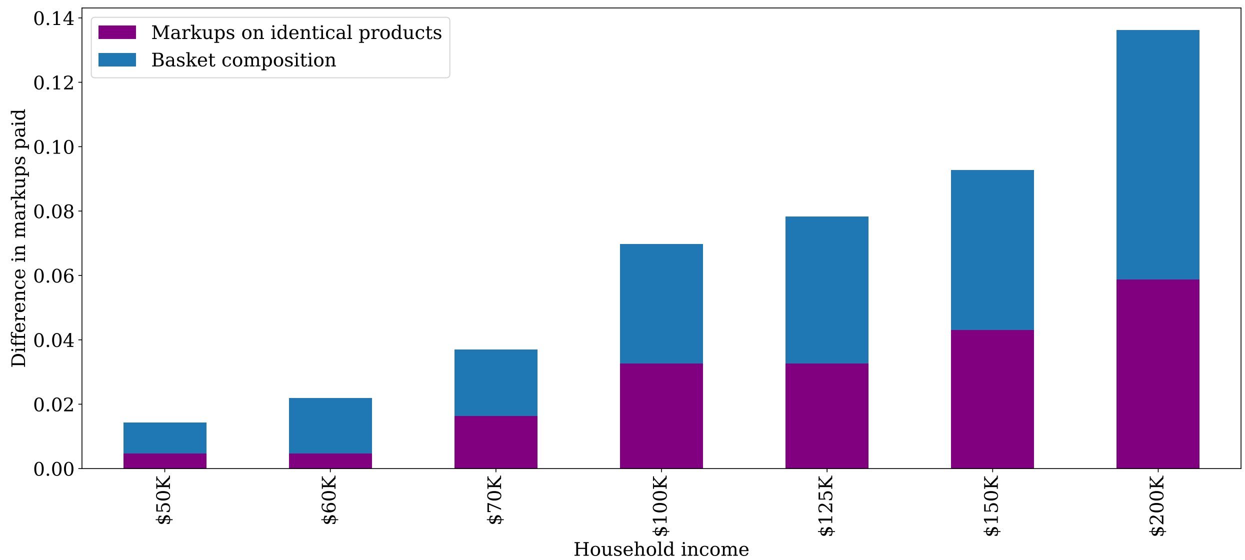
	Measure of wholesale cost	
	Base Price	Deal Price
Percentiles of distribution:		
10	1.053	1.119
25	1.204	1.288
50	1.382	1.470
75	1.600	1.694
90	1.911	2.002
Percent below $\mu = 1$:		
By count	6.96	4.72
By sales	12.63	6.35
No. UPCs matched	67161	67161

Table A3: Coverage of UPC wholesale costs data by income level.

	Percent matched to wholesale cost data		Average relative unit price ($\hat{p}$)	
Income group	Transactions	Expenditures	Matched	Unmatched
$10–25K	41	38	-0.02	-0.05
$25–40K	42	38	0.00	-0.02
$40–60K	43	38	0.04	0.02
$60–100K	44	37	0.09	0.09
Over $100K	44	35	0.17	0.17
All	43	37	0.06	0.05

Appendix B Additional Tables and Figures

Figure B1: Decomposition of the difference in markups paid between each income group and the group of households with below \$50K in income.

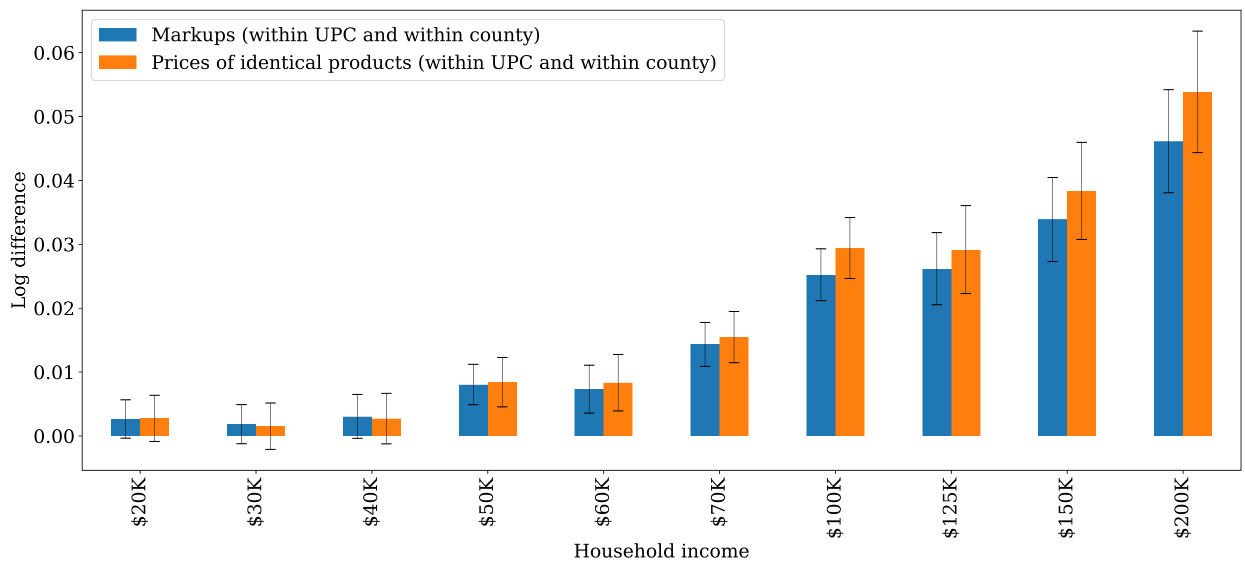


Note: The figure decomposes the difference in average markups paid by income group i relative to the group with below \$50K in income. The difference between the sales-weighted average markup paid by group i and the reference group o is decomposed as

$$\begin{aligned}\mu_i^{\text{avg}} - \mu_o^{\text{avg}} &= \sum_k \lambda_{ik}\mu_{ik} - \sum_k \lambda_{ok}\mu_{ok} \\ &= \underbrace{\sum_{k \in \mathcal{I} \cup \mathcal{O}} \lambda_{ok}(\mu_{ik} - \mu_{ok})}_{\substack{\text{Diff. markups paid} \\ \text{for identical products}}} + \underbrace{\sum_{k \in \mathcal{I} \cup \mathcal{O}} \mu_{ik}(\lambda_{ik} - \lambda_{ok}) + \left(\sum_{k \in \mathcal{I} \setminus \mathcal{O}} \lambda_{ik}\mu_{ik} - \sum_{k \in \mathcal{O} \setminus \mathcal{I}} \lambda_{ok}\mu_{ok} \right)}_{\substack{\text{Differences in} \\ \text{basket composition}}}\end{aligned}$$

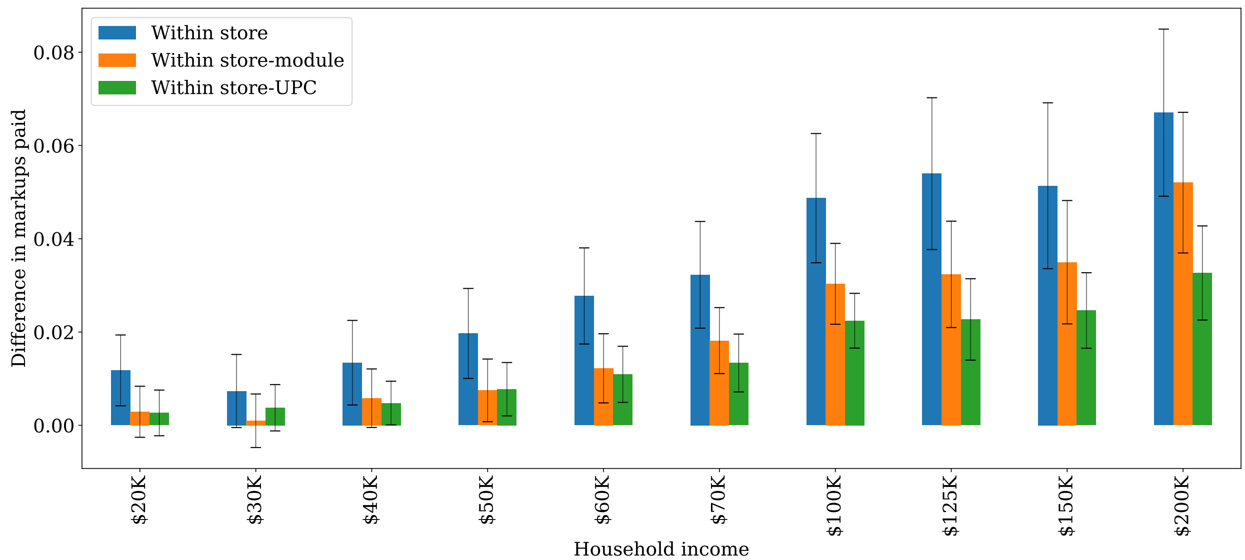
where $\mathcal{I}$ and $\mathcal{O}$ are the sets of UPCs purchased by groups i and o, μ_{ik} and μ_{ok} are the sales-weighted average markup paid by groups i and o for UPC k, and λ_{ik} and λ_{ok} are the share of group i and o's total expenditures spent on UPC k.

Figure B2: Difference in log markups paid for identical products within county (blue) and difference in log prices paid for identical products within county (orange), relative to households with below $20K income.



Note: The blue bars plot the coefficients β_ℓ from a sales-weighted regression of log markup paid on household income dummies, demographic controls, county fixed effects, and UPC fixed effects. The orange bars plot the coefficients β_ℓ from an analogous regression (specification (4)), where the dependent variable is instead log price paid. Standard errors are two-way clustered by product brand and household county.

Figure B3: Difference in markups paid within store (blue), within store-product module (orange), and within store-UPC (green) relative to households with below \$20K income.

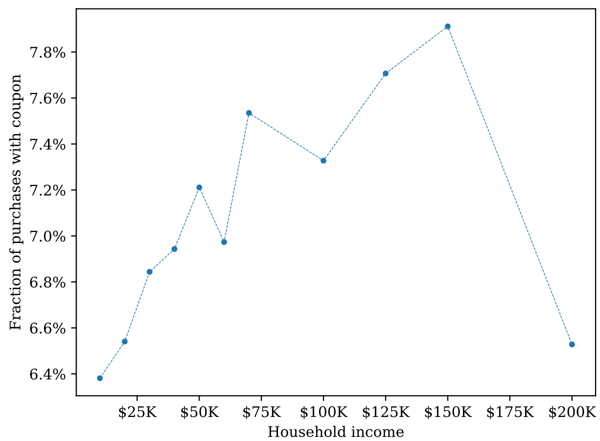


Note: This figure plots the coefficients β_ℓ on household income dummies in a sales-weighted regression of the markup paid by a household on demographic controls (race, ethnicity, household size, and presence and age of female head of household) and store fixed effects. The orange bars add store-product module fixed effects, and the green bars add store-UPC fixed effects:

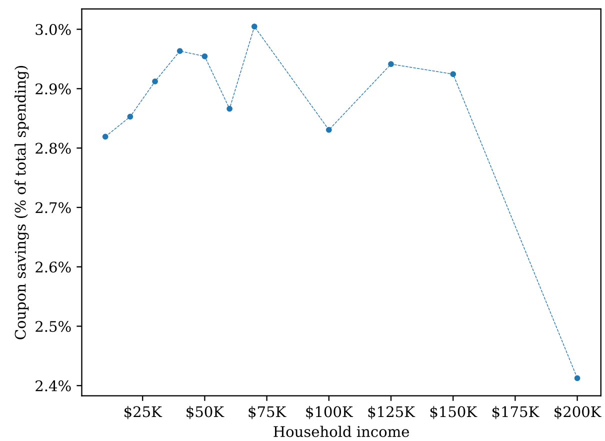
$$\text{Markup}_{i,g} = \sum_{\ell} \beta_\ell 1\{i \text{ has income level } \ell\} + \gamma' X_i + \alpha_{\text{Store}} + \tilde{\alpha}_{\text{Store-Module}} + \hat{\alpha}_{\text{Store-UPC}} + \epsilon_{i,g}.$$

Income levels on the y-axis are the minimum of the income bracket provided by Nielsen (e.g., \$30K includes households reporting income between \$30–40K). Standard errors are two-way clustered by product brand and household county.

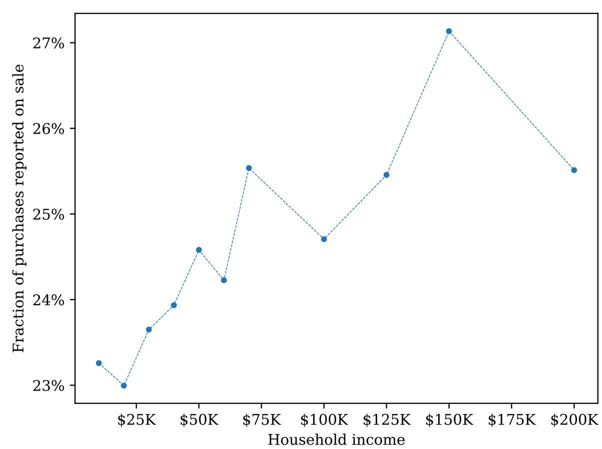
Figure B4: Savings technologies: Coupons and sales.



(a) Fraction of purchases made with a coupon.

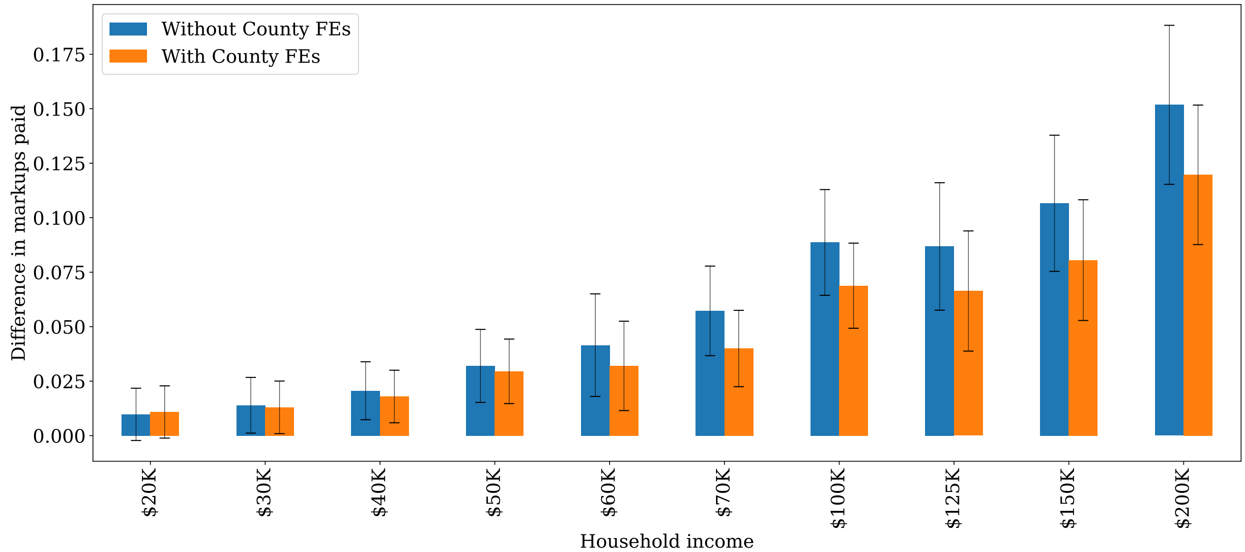


(b) Savings due to coupon use (as a percent of total spending).

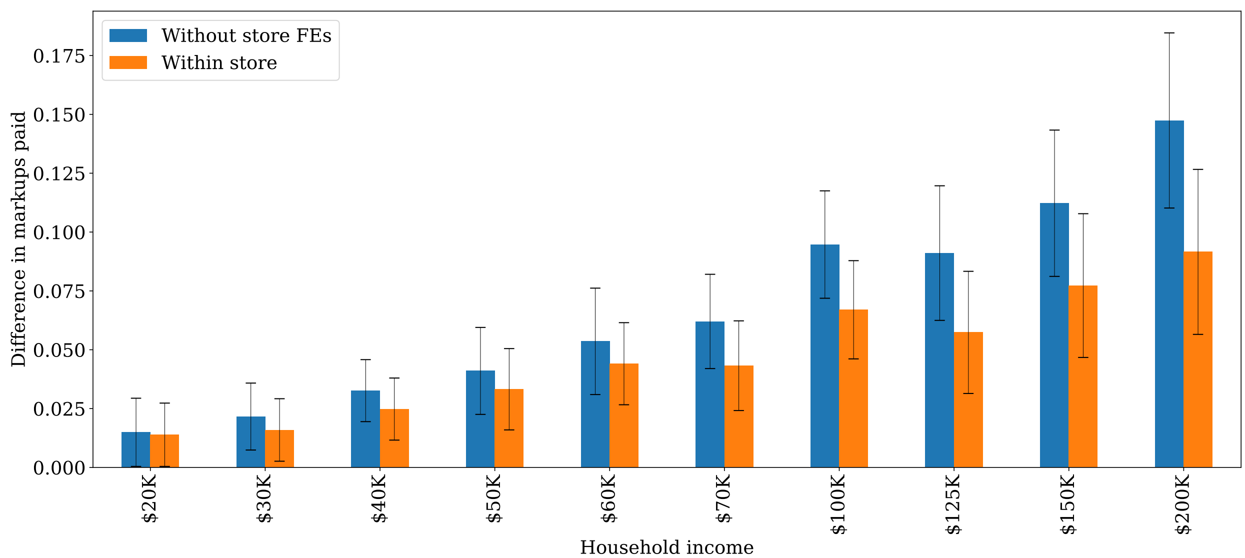


(c) Fraction of purchases self-reported as on sale.

Figure B5: Difference in markups paid relative to households with below \$20K income, using PromoData wholesale prices matched to household purchases by ScanTrack market.



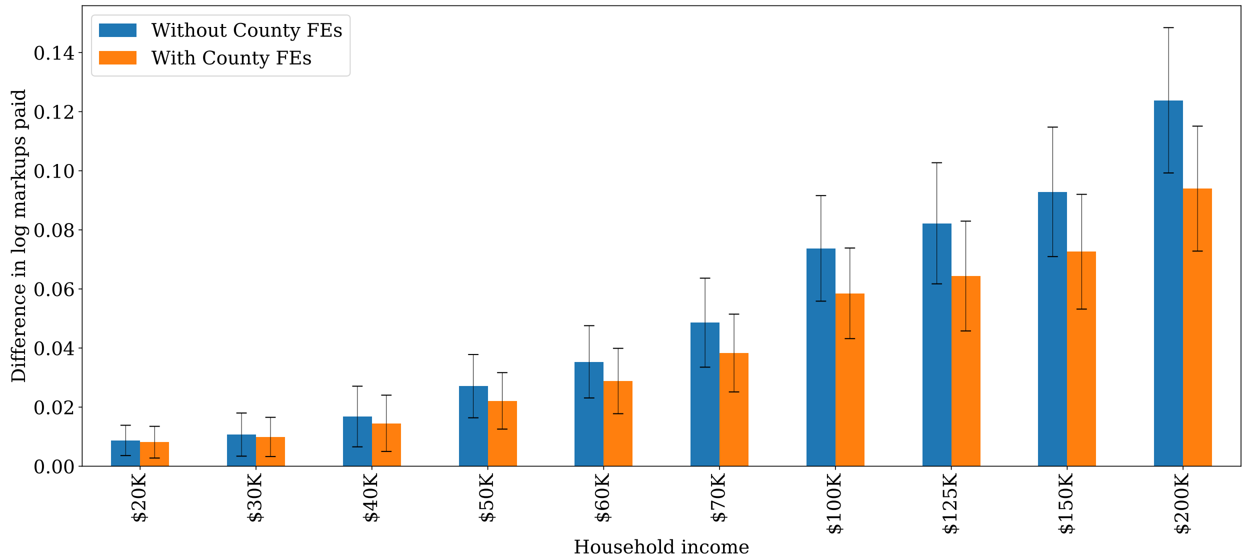
(a) With and without county fixed effects (N = 3.0 million).



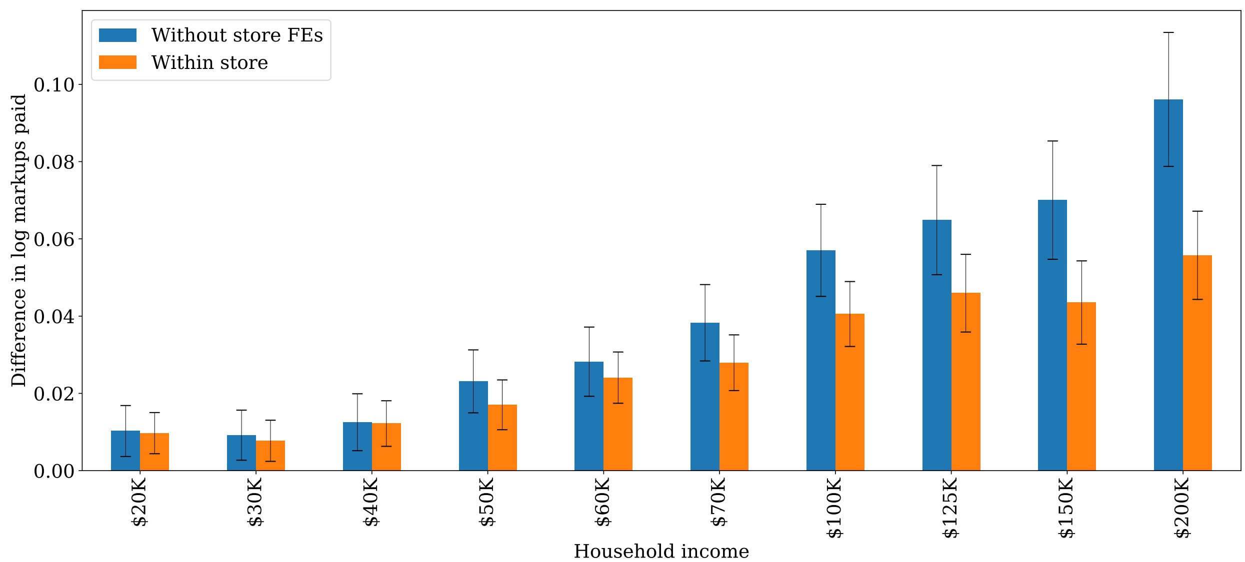
(b) With and without store fixed effects (store transactions only, N = 1.8 million).

Note: These figures plot the coefficients β_ℓ on household income dummies in a sales-weighted regression of the markup paid by a household on demographic controls (race, ethnicity, household size, and age of female head of household). Income levels on the y-axis are the minimum of the income bracket provided by Nielsen (e.g., \$30K includes households reporting income between \$30–40K). Standard errors are two-way clustered by product brand and household county. Figure (a) shows β_ℓ with and without county fixed effects (specification (1)), and (b) shows $\tilde{\beta}_\ell$ with and without store fixed effects for the sample of transactions where the unique store ID is observed (specification (2)).

Figure B6: Difference in log markups paid relative to households with below $20K income.



(a) With and without county fixed effects (N = 25.8 million).



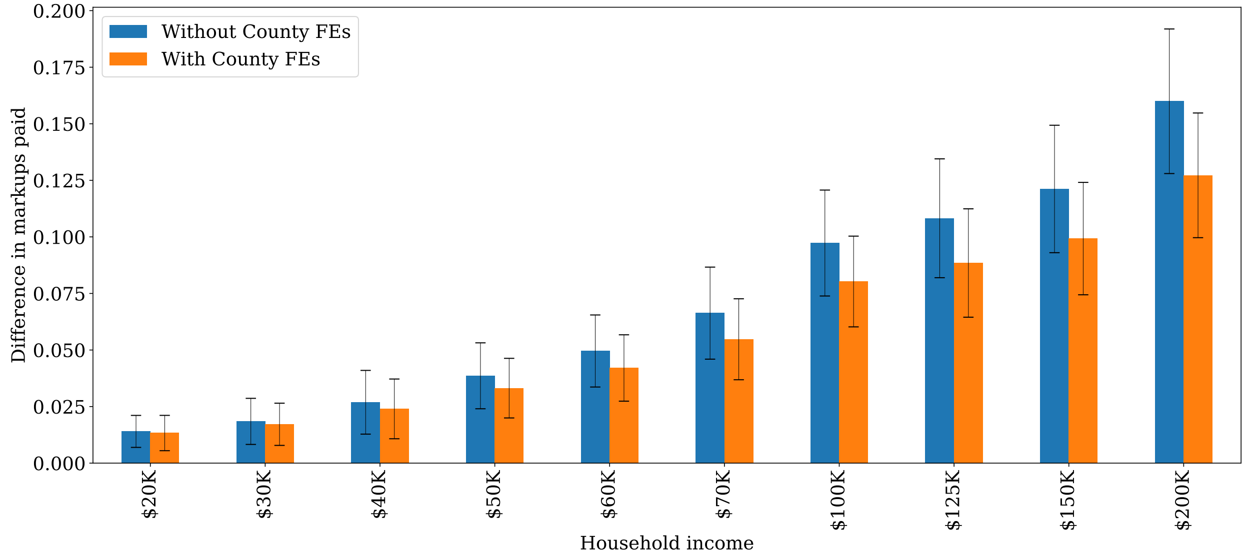
(b) With and without store fixed effects (store transactions only, N = 14.0 million).

Note: These figures plot the coefficients β_ℓ on household income dummies in a sales-weighted regression of the log markup paid by a household on demographic controls. Income levels on the y-axis are the minimum of the income bracket provided by Nielsen. Standard errors are two-way clustered by product brand and household county. Figure (a) shows β_ℓ with and without county fixed effects, and (b) shows $\tilde{\beta}_\ell$ with and without store fixed effects for the sample of transactions where the unique store ID is observed.

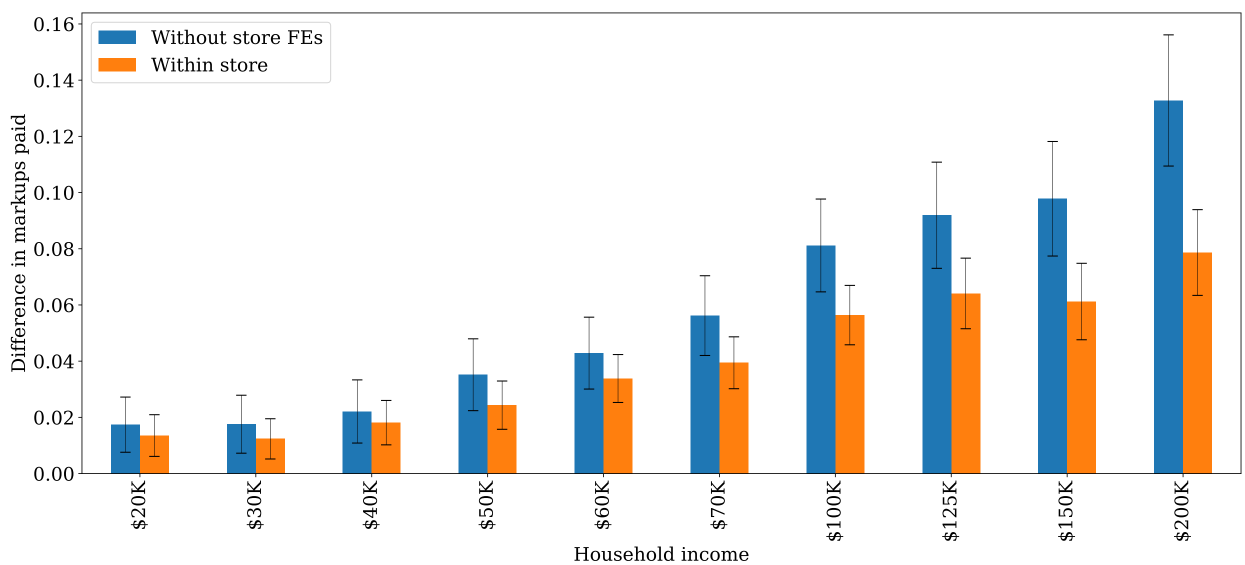
$$\text{(a)} \qquad \log(\text{Markup})_{i,g} = \sum_\ell \beta_\ell 1\{i \text{ has income level } \ell\} + \gamma' X_i + \delta_{\text{County}} + \epsilon_{i,g}.$$

$$\text{(b)} \qquad \log(\text{Markup})_{i,g} = \sum_\ell \tilde{\beta}_\ell 1\{i \text{ has income level } \ell\} + \tilde{\gamma}' X_i + \alpha_{\text{Store}} + \epsilon_{i,g}.$$

Figure B7: Difference in markups paid relative to households with below $20K income, weighting regressions by costs.



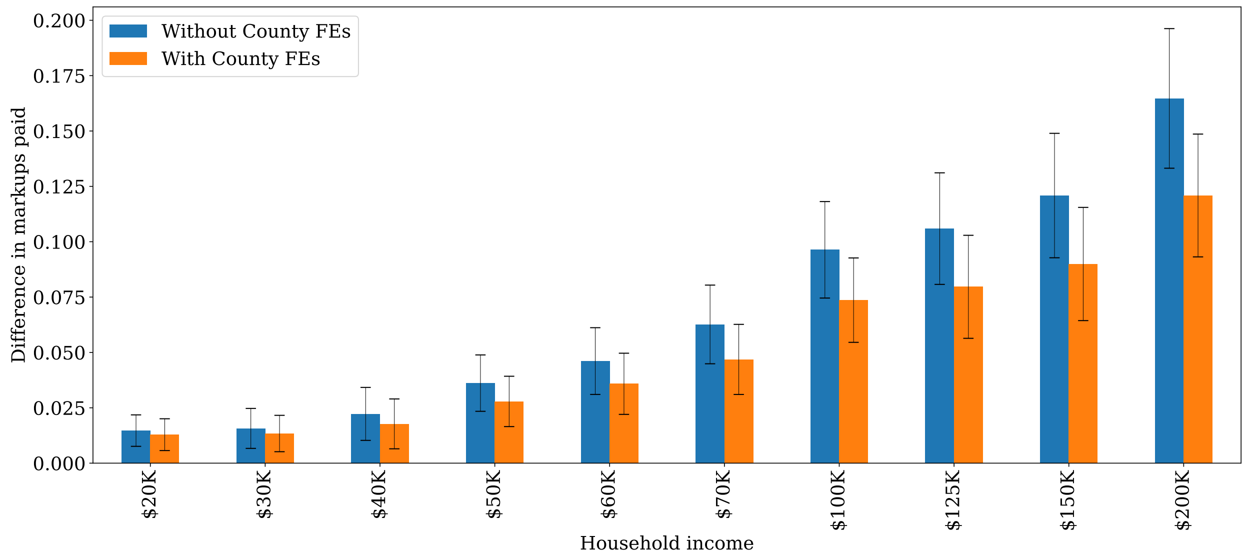
(a) With and without county fixed effects (N = 25.8 million).



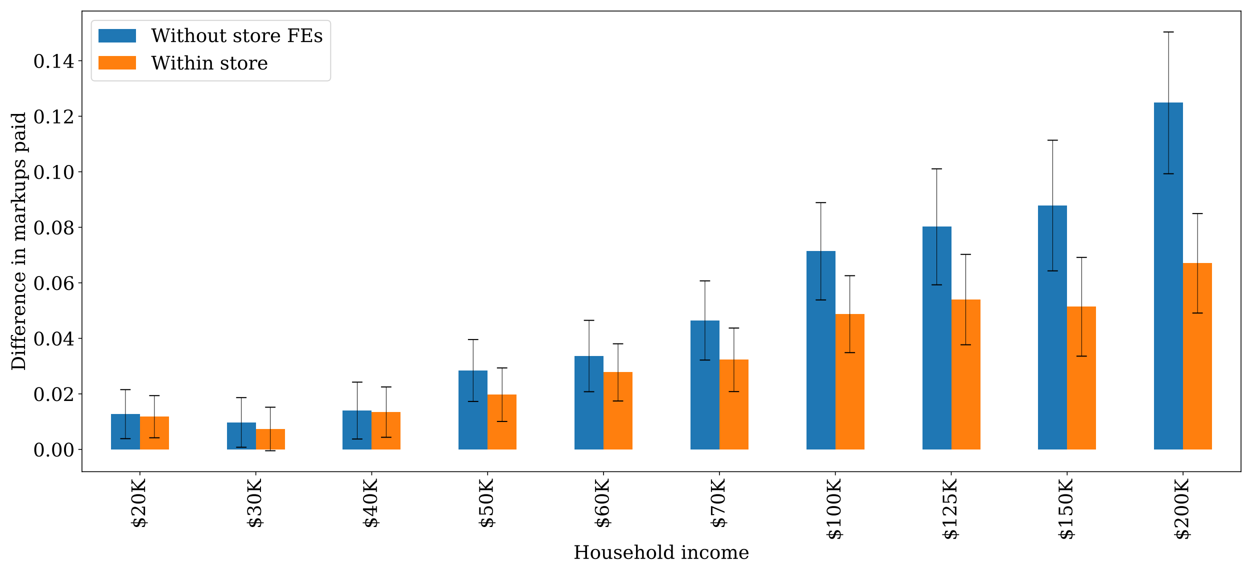
(b) With and without store fixed effects (store transactions only, N = 14.0 million).

Note: These figures plot the coefficients β_ℓ on household income dummies in a regression of the markup paid by a household on demographic controls weighted by costs (sales / markup). Income levels on the y-axis are the minimum of the income bracket provided by Nielsen. Standard errors are two-way clustered by product brand and household county. Figure (a) shows β_ℓ with and without county fixed effects (specification (1)), and (b) shows $\tilde{\beta}_\ell$ with and without store fixed effects for the sample of transactions where the unique store ID is observed (specification (2)).

Figure B8: Difference in markups paid relative to households with below \$20K income (using Price-Trak deal prices as measure of wholesale costs).



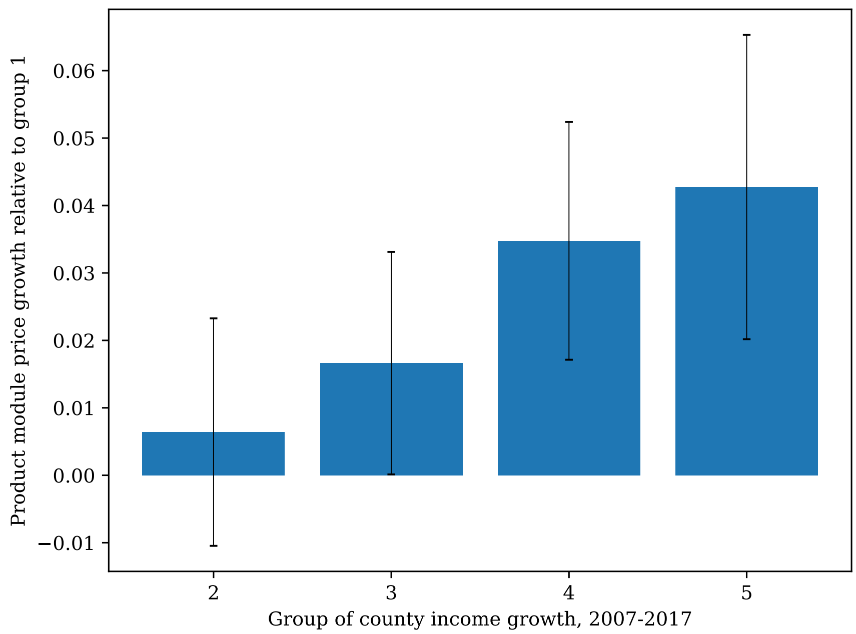
(a) With and without county fixed effects (N = 25.8 million).



(b) With and without store fixed effects (store transactions only, N = 14.0 million).

Note: These figures plot the coefficients β_ℓ on household income dummies in a sales-weighted regression of the markup paid by a household on demographic controls (race, ethnicity, household size, and age of female head of household). Income levels on the y-axis are the minimum of the income bracket provided by Nielsen (e.g., \$30K includes households reporting income between \$30–40K). Standard errors are two-way clustered by product brand and household county. Figure (a) shows β_ℓ with and without county fixed effects (specification (1)), and (b) shows $\tilde{\beta}_\ell$ with and without store fixed effects for the sample of transactions where the unique store ID is observed (specification (2)).

Figure B9: Product module price growth against quintile of county income growth, 2007–2017.

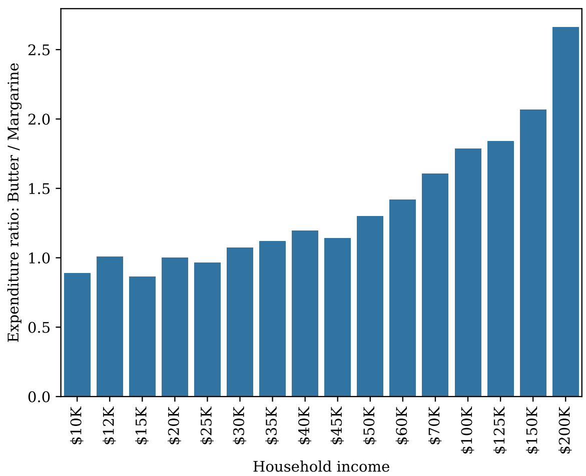


Note: The figure plots fixed effects β_i estimated from the regression

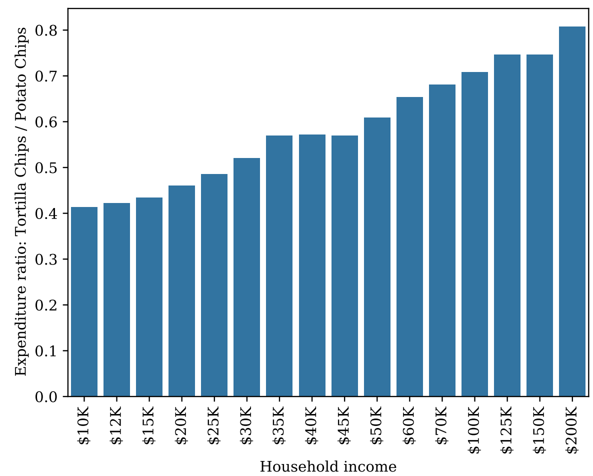
$$\frac{\text{Avg. Unit Price}_{m,c,2017}}{\text{Avg. Unit Price}_{m,c,2007}} = \sum_{i=1}^{5} \beta_i 1\left\{\text{County } c \text{ is in quintile } i \text{ of income growth}\right\} + \alpha_m + \epsilon_{m,c},$$

where m denotes a product module, c denotes a county, α_m are product module fixed effects, and $\epsilon_{m,c}$ is a mean-zero error. The indicator for the first quintile of county income growth is excluded, so that the fixed effects plotted are relative to the group of counties with lowest per-capita income growth from 2007 to 2017.

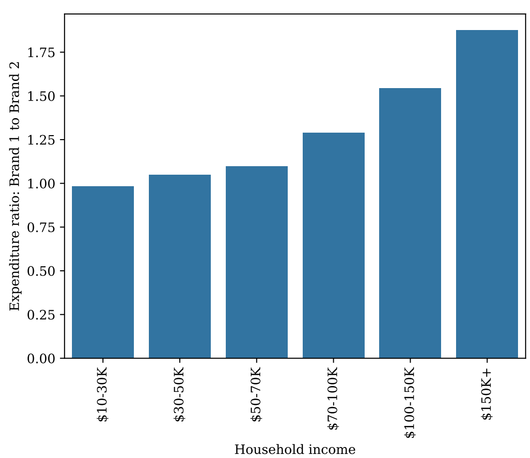
Figure B10: Examples of substitution across product modules and brands.



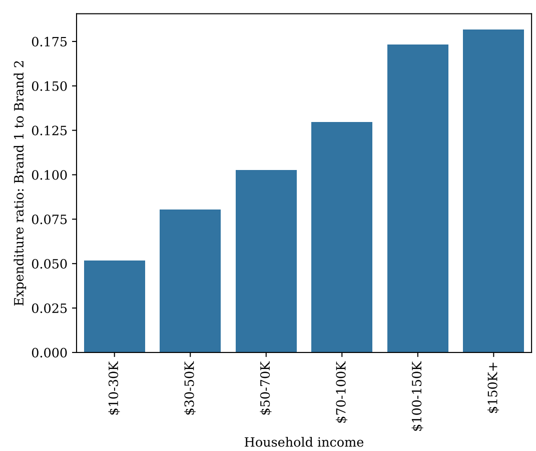
(a) Expenditures on butter (avg. markup 45%) vs. margarine (avg. markup 33%).



(b) Expenditures on tortilla chips (avg. markup 50%) vs. potato chips (avg. markup 19%).



(c) Expenditures on margarine brand 1 (avg. markup 38%) vs. brand 2 (avg. markup 23%).



(d) Expenditures on potato chips brand 1 (avg. markup 33%) vs. brand 2 (avg. markup 18%).

Note: Panels (a) and (b) show the ratio of expenditures on two product modules: (a) butter versus margarine and (b) tortilla chips versus potato chips by household income group. The sales-weighted average of markups on butter UPCs is 45% compared to 33% for margarine UPCs, and 50% for tortilla chips compared to 19% for potato chips. Panels (c) and (d) show the ratio of expenditures on (c) two margarine brands and (d) two potato chip brands. The average markup of margarine brand 1 is 38% compared to 23% for margarine brand 2, and the average markup of potato chip brand 1 is 33% compared to 18% of potato chip brand 2.

Figure B11: Density $dH(i)$, constructed from data by Saez and Zucman (2019).

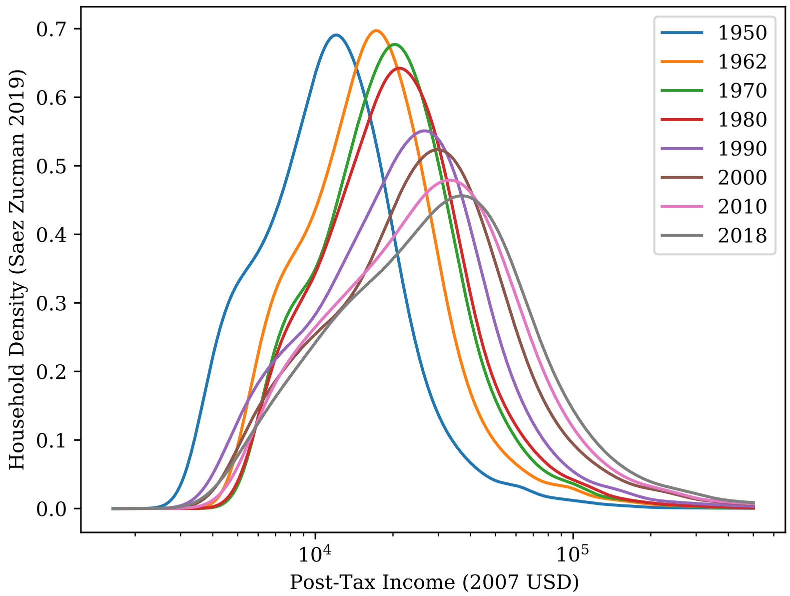
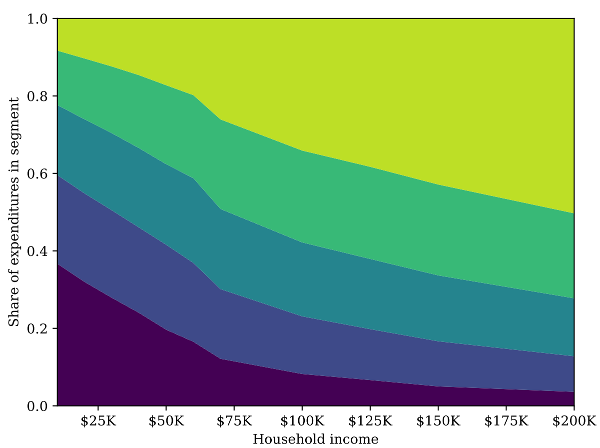
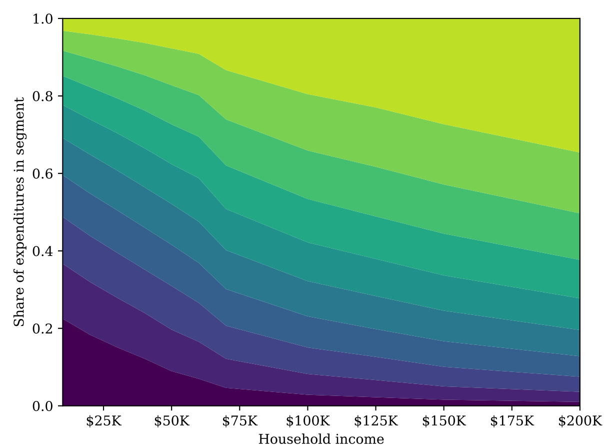


Figure B12: Segmentation of UPCs by household income.

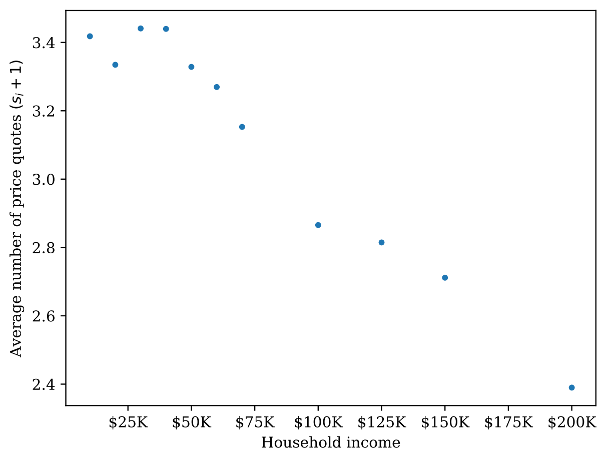


(a) Share of purchases in each segment: $K = 5$.

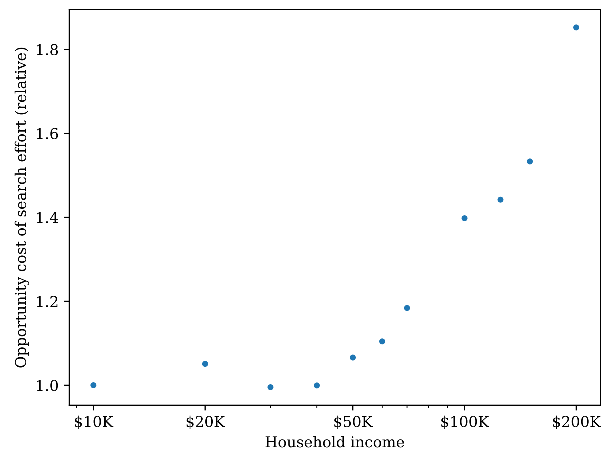


(b) Share of purchases in each segment: $K = 10$.

Figure B13: Price quotes received and opportunity cost of search effort, for $K = 10$.

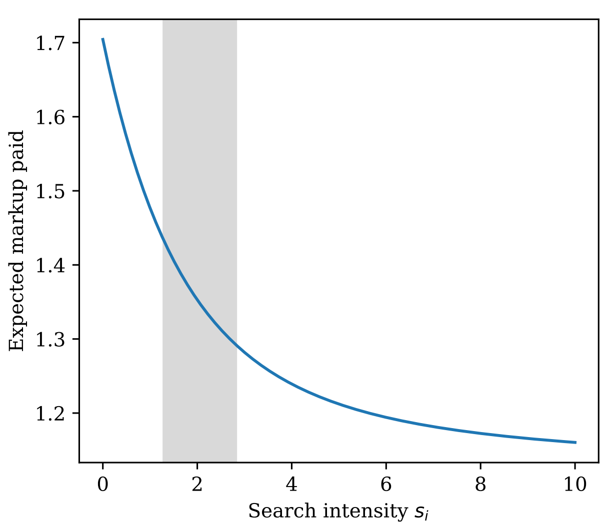


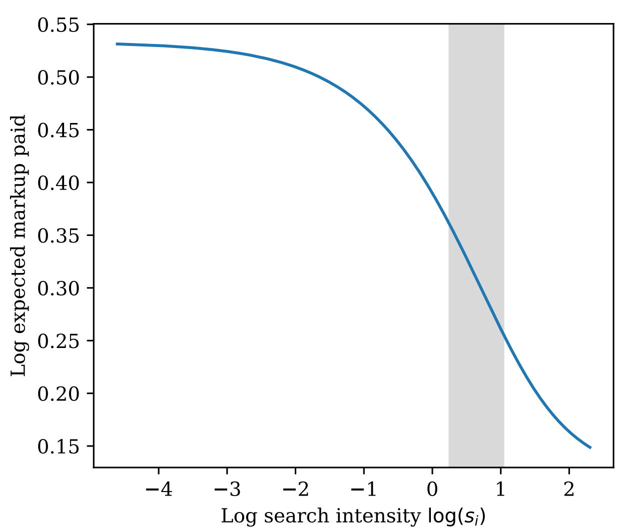
(a) Average no. of price quotes received ($s(z) + 1$).



(b) Oppty. cost $\phi(z)$ (normalized to one for <$20K).

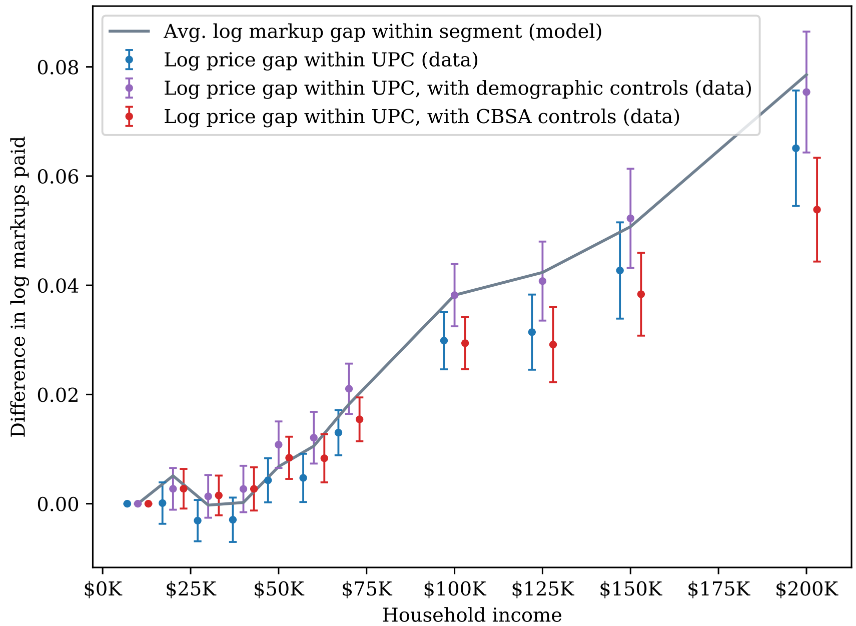
Figure B14: Returns to consumer search.





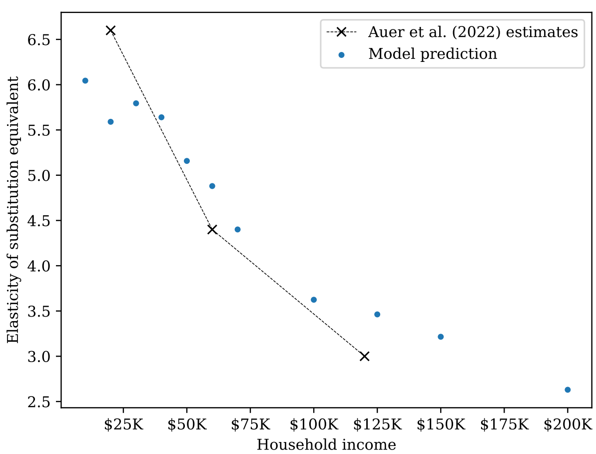
Note: The shaded region indicates the range of search intensities s_i chosen by income groups in the baseline calibration ($K = 1$). Doubling search intensity in the shaded region results in a 7.5–9.2% decrease in prices paid.

Figure B15: Difference in log markups paid within segment in the model with $K = 10$ segments, compared to estimates of difference in log prices paid for identical products in the data.

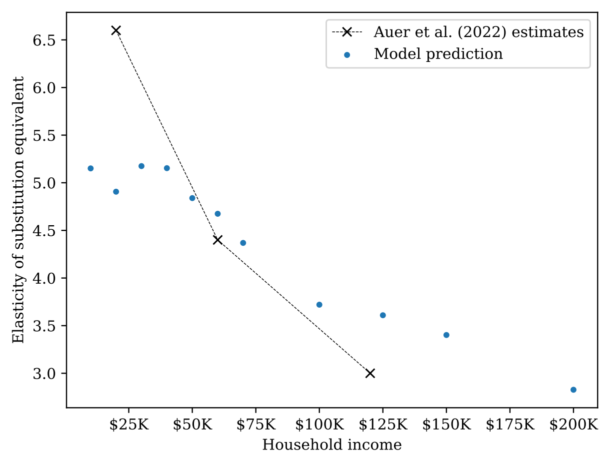


Note: The grey line shows the average log markup gap within segment in the model calibrated with $K = 10$ segments. The scatter plots and error bars show fixed effects on income level in a regression of log prices on household income level and UPC fixed effects.

Figure B16: Comparison of price sensitivity in calibrated model to estimates of elasticity of substitution from Auer et al. (2022).



(a) Single-product model ($K = 1$).



(b) Model with segmentation ($K = 10$).

Figure B17: Predicted search intensity by real income level (in 2007 dollars) from 1950–2018.

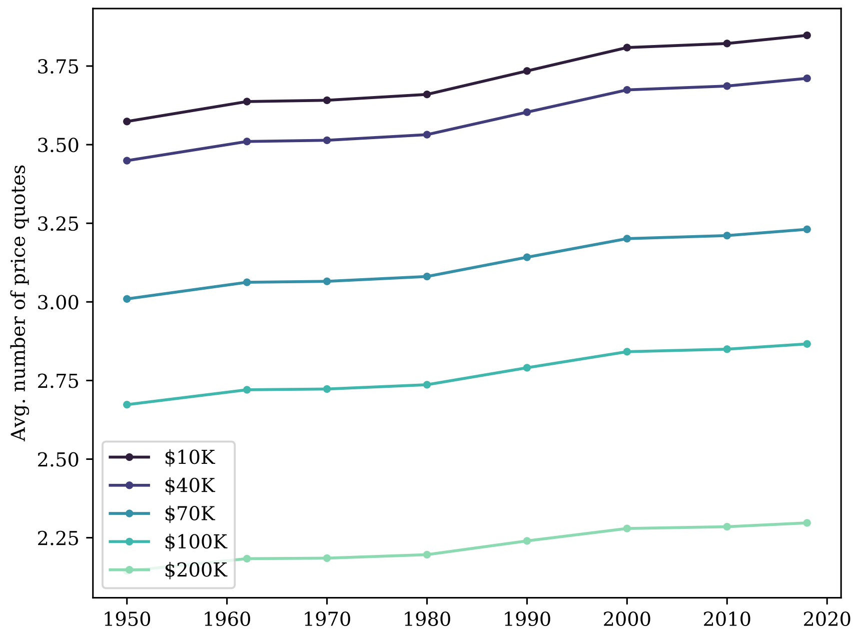
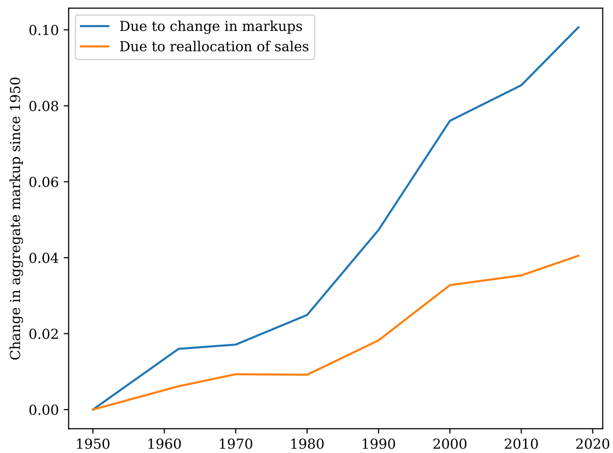
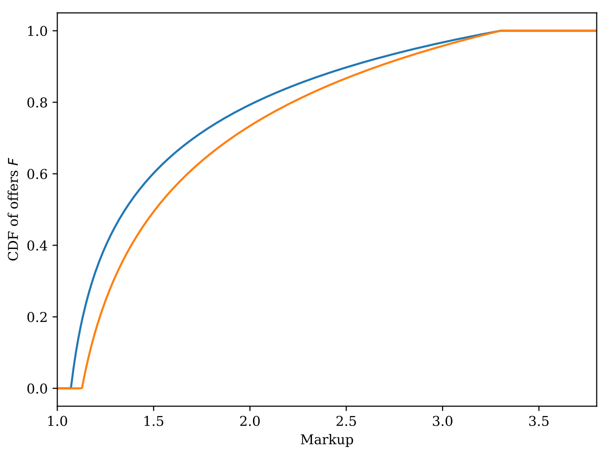


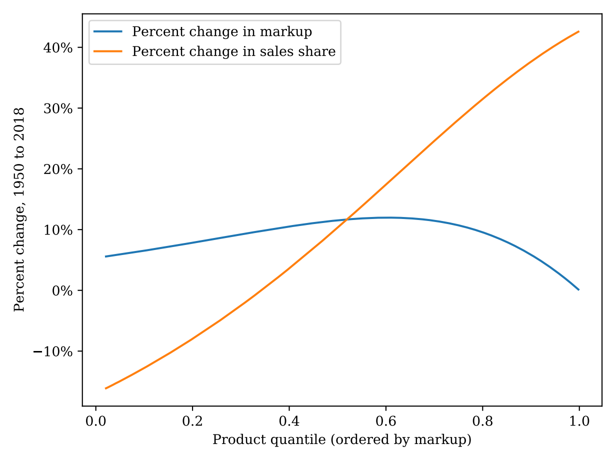
Figure B18: Role of reallocations and change in the offer price distribution, 1950–2018.



(a) Decomposition of change in agg. markup.

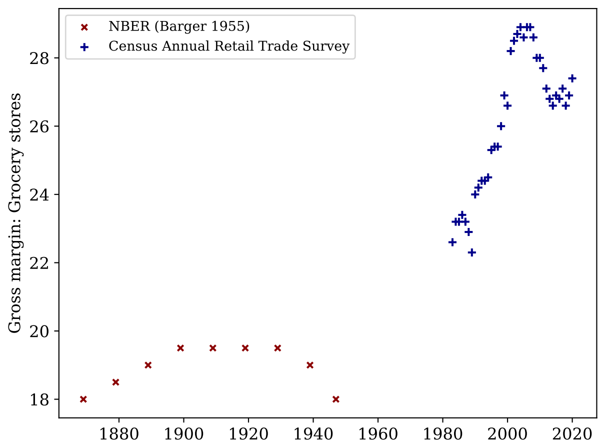


(b) Offer distribution F in 1950 and 2018.

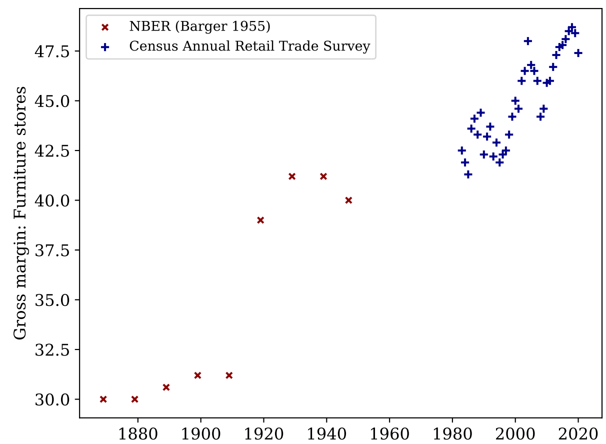


(c) Percent change in markups and sales shares by quantile from 1950 and 2018.

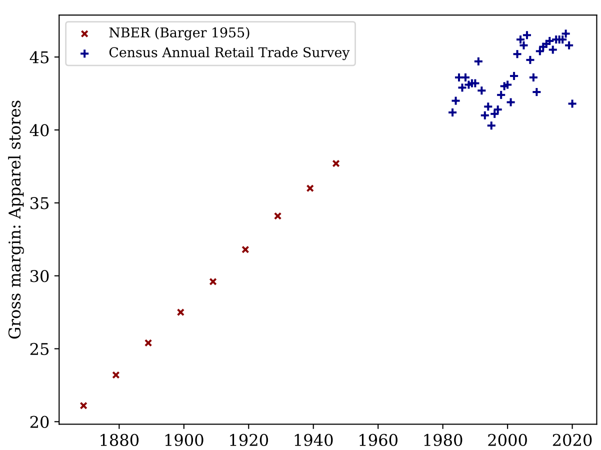
Figure B19: Data on retail gross margins over time by subsector.



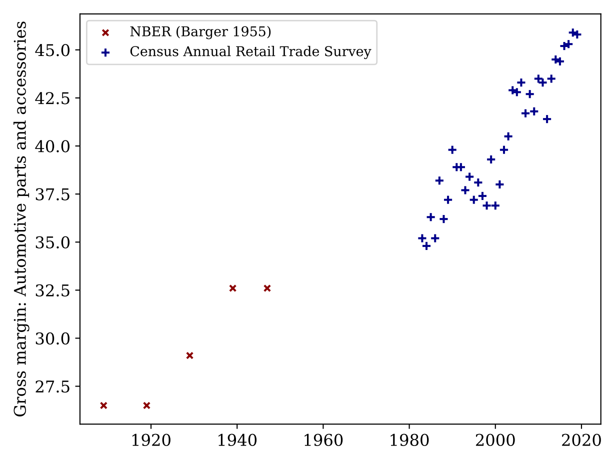
(a) Grocery stores



(b) Furniture stores



(c) Apparel stores



(d) Automobile accessory stores

Note: Gross margin estimates are available for selected years from 1869 to 1947 from Barger (1955), and annually from the Census Annual Retail Trade Survey from 1983 to 2020. Both sources report gross margins as total sales less total costs of goods sold as a percent of total sales. Grocery stores include SIC 541 from 1983–1992 and NAICS 4451 after 1993. Furniture stores include SIC 571 from 1983–1992 and NAICS 442 after 1993. Apparel stores include SIC 56 from 1983–1992 and NAICS 448 after 1993. Automobile parts and accessory stores include SIC 553 from 1983–1992 and NAICS 4413 after 1993.

Table B1: Share of markup gap within county and within store for alternative measures and other years from 2006–2009.

	Percent of markup gap within county (%)			Percent of markup gap within store (%)		
Income group	\$70–100K	\$100–125K	\$200K+	\$70–100K	\$100–125K	\$200K+
Baseline (2007 data)	74.8	76.4	73.4	66.6	66.1	53.1
Cost-weighted markups	82.6	82.5	79.5	70.1	69.5	59.2
Log markups	78.5	79.1	75.9	69.5	68.7	57.0
Using PromoData deal price	72.0	74.3	71.5	59.8	60.9	47.7
Using PromoData market data	69.8	77.6	78.8	69.7	70.7	62.1
Without demographic controls	84.5	77.1	73.8	77.4	69.8	52.4
Only top 100 product modules	74.3	76.2	73.3	71.5	70.3	60.0
Excluding perishable items	71.3	73.9	71.4	60.3	62.5	49.7
2006 data	73.1	75.6	75.1	74.6	70.0	57.5
2008 data	80.2	77.1	72.0	81.1	71.4	52.5
2009 data	78.6	78.2	71.1	77.7	66.8	55.0

Table B2: Comparison of within-store elasticity of markups paid to income, using national wholesale price versus market-level wholesale price.

	National price		Market price	
	Base	Deal	Base	Deal
	(1)	(2)	(3)	(4)
Log Household Income	0.020**	0.017**	0.025**	0.023**
	(0.002)	(0.002)	(0.003)	(0.003)
Household Size	−0.006**	−0.003**	−0.007**	−0.005**
	(0.001)	(0.001)	(0.002)	(0.002)
Demographic Controls	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Store FEs	Yes	Yes	Yes	Yes
N (millions)	14.0	14.0	1.81	1.81
R^2	0.08	0.09	0.09	0.11

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $13.5K). Demographic controls include household size, race, ethnicity, presence of female head of household, and age group of female head of household. Regression weighted by sales, and standard errors two-way clustered by UPC and household. * indicates significance at 10%, ** at 5%.

Table B3: Impact of income on markups paid (using Price-Trak deal prices as measure of wholesale costs).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Household Income	0.032**	0.038**	0.032**	0.029**	0.021**	0.017**	0.012**	0.011**	0.009**
	(0.004)	(0.005)	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Log Avg. CBSA Income			0.102**						
			(0.011)						
Household Size		−0.003**	−0.003**	−0.002**	−0.003**	−0.003**	−0.003**	−0.003**	−0.003**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Demographic Controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FEs				Yes	Yes	Yes	Yes	Yes	Yes
Store FEs						Yes	Yes	Yes	Yes
Store-Group FEs							Yes		
Store-Module FEs								Yes	
Store-UPC FEs									Yes
N (millions)	25.8	25.8	23.8	25.8	14.0	14.0	14.0	14.0	14.0
R^2	0.00	0.00	0.01	0.02	0.03	0.09	0.40	0.61	0.92

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $13.5K). Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

Table B4: Impact of income on markups paid, instrumenting for household income.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Household Income	0.035**	0.054**	0.051**	0.049**	0.046**	0.039**	0.030**	0.025**	0.017**
	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
Log Avg. CBSA Income			0.093**						
			(0.010)						
Household Size		−0.008**	−0.008**	−0.007**	−0.009**	−0.008**	−0.006**	−0.006**	−0.004**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Demographic Controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FEs				Yes	Yes	Yes	Yes	Yes	Yes
Store FEs						Yes	Yes	Yes	Yes
Store-Group FEs							Yes		
Store-Module FEs								Yes	
Store-UPC FEs									Yes
N (millions)	25.8	25.8	23.8	25.8	14.0	14.0	14.0	14.0	14.0
R^2	0.00	0.00	0.01	0.02	0.02	0.08	0.42	0.63	0.92

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of \$13K would be in the range \$12–15K and would be assigned a value of \$13.5K). Log household income is instrumented with the log of total observed household expenditures and fixed effects for education, employment status, and occupation group for both the male and female heads of household. Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

Table B5: Impact of income on prices paid.

	(1)	(2)	(3)	(4)	(5)	(6)
Log Household Income	0.017**	0.020**	0.016**	0.015**	0.017**	0.015**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log Avg. CBSA Income			0.090**			
			(0.008)			
Household Size		−0.004**	−0.004**	−0.004**	−0.004**	−0.003**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
UPC FEs	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls		Yes	Yes	Yes	Yes	Yes
County FEs				Yes	Yes	Yes
Store FEs						Yes
N (millions)	25.8	25.8	23.8	25.8	14.0	14.0
R^2	0.96	0.96	0.96	0.96	0.96	0.96

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $13.5K). Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

Table B6: Relationship between buyer income and UPC retail markup.

Log Retail Markup	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Household Income	0.033**	0.033**	0.030**	0.032**	0.020**	0.020**	0.020**	0.021**	0.019**	0.020**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log Avg. CBSA Income	0.104**	0.104**	0.106**	0.106**	0.100**	0.099**	0.100**	0.100**	0.101**	0.100**
	(0.010)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Necessity		−0.043				−0.151**				
		(0.042)				(0.024)				
Log Unique UPCs in Module			−0.107**				−0.061**			
			(0.017)				(0.008)			
Log Unique Brands in Module			0.085**				0.047**			
			(0.016)				(0.012)			
UPC Sales Share				0.343**				0.134		−0.014
				(0.157)				(0.085)		(0.157)
Brand Sales Share				0.246**				0.036		0.041
				(0.055)				(0.044)		(0.044)
Firm Sales Share				−0.153**				−0.094**		−0.077*
				(0.055)				(0.042)		(0.042)
Module HHI				−0.094				0.088		
				(0.075)				(0.057)		
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Group FEs					Yes	Yes	Yes	Yes		
Product Module FEs									Yes	Yes
N (millions)	23.8	23.8	23.8	23.8	23.8	23.8	23.8	23.8	23.8	23.8
R^2	0.01	0.01	0.04	0.03	0.19	0.19	0.19	0.19	0.29	0.29

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $13.5K). Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. CBSA income is from the BEA. Necessity is measured as the share of households in the Homescan panel that purchase at least one good from that module in the year. UPC sales share, brand sales share, and the product module Herfindahl-Hirschman index are all calculated using sales in the Homescan data. * indicates significance at 10%, ** at 5%.

Table B7: Relationship between buyer income and UPC retail markup.

Log Retail Markup	(1)	(2)	(3)	(4)
Log Household Income	0.016**	0.016**	0.015**	0.015**
	(0.001)	(0.001)	(0.001)	(0.001)
Log Avg. Income: Other UPC Buyers	0.189**	0.212**	0.192**	0.205**
	(0.046)	(0.046)	(0.044)	(0.043)
Necessity		−0.154**		
		(0.027)		
Log Unique UPCs in Module			−0.090**	
			(0.014)	
Log Unique Brands in Module			0.069**	
			(0.015)	
UPC Sales Share				0.377**
				(0.150)
Brand Sales Share				0.222**
				(0.054)
Firm Sales Share				−0.167**
				(0.045)
Module HHI				−0.037
				(0.066)
Demographic Controls	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Store FEs	Yes	Yes	Yes	Yes
N (millions)	14.0	14.0	14.0	14.0
R^2	0.09	0.10	0.11	0.10

Note: The log of household income is calculated using the midpoint of the range provided by Nielsen (e.g., a household with an income of \$13K would be in the range \$12–15K and would be assigned a value of \$13.5K). Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Necessity is measured as the share of households in the Homescan panel that purchase at least one good from that module in the year. UPC sales share, brand sales share, and the product module Herfindahl-Hirschman index are all calculated using sales in the Homescan data. * indicates significance at 10%, ** at 5%.

Table B8: Elasticity of unit prices to CBSA and county income across space and over time, 2004 and 2019.

Log Unit Price	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Household Income	0.086**	0.089**	0.087**	0.088**	0.087**	0.086**	0.089**	0.084**	0.085**	0.084**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log Avg. CBSA Income	0.209*	0.174**	0.323	0.040**	0.029**					
	(0.023)	(0.009)	(0.056)	(0.0001)	(0.002)					
Log Avg. County Income						0.179**	0.152**	0.317	0.059	0.051
						(0.013)	(0.006)	(0.078)	(0.020)	(0.020)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product Module FEs	Yes					Yes				
Year-Product Module FEs		Yes					Yes			
CBSA-Product Module FEs			Yes							
Year-Product Module FEs and CBSA FEs				Yes						
CBSA-Product Module FEs and Year FEs					Yes					
County-Product Module FEs								Yes		
Year-Product Module FEs and County FEs									Yes	
County-Product Module FEs and Year FEs										Yes
N (millions)	83.5	83.5	83.5	83.5	83.5	88.7	88.7	88.7	88.7	88.7
R^2	0.86	0.86	0.87	0.86	0.87	0.86	0.86	0.88	0.86	0.88

Note: Sample includes all Homescan purchases in 2004 and 2019. For the purpose of this analysis, "Product Modules" are defined as product modules from Nielsen further disaggregated by unit of account. For example, all butter products with units in ounces are counted as a module, and distinct from butter products with units in pounds. All unit prices, incomes, and sales are deflated to 2007 USD using the GDP deflator from FRED, and regressions are weighted by (deflated) sales. Standard errors two-way clustered by brand and county. * indicates significance at 10%, ** at 5%.

Table B9: Robustness: Effect of income and county income on search intensity.

Dependent variable (Measure of search intensity)	Log(Income) Coefficient	Log(Income) SE	Log(Avg. County Income) Coefficient	Log(Avg. County Income) SE	County FEs	State FEs	Control for Log(Grocery Estabs.)
OLS estimates:							
Log(Shopping trips per $1K spent)	-0.25**	(0.01)			Yes	No	No
Log(Shopping trips per $1K spent)	-0.26**	(0.01)	0.03	(0.02)	No	Yes	Yes
Log(Shopping trips per transaction)	-0.11**	(0.00)			Yes	No	No
Log(Shopping trips per transaction)	-0.12**	(0.00)	0.11**	(0.03)	No	Yes	Yes
Log(Shopping trips per brand bought)	-0.11**	(0.00)			Yes	No	No
Log(Shopping trips per brand bought)	-0.12**	(0.00)	0.10**	(0.03)	No	Yes	Yes
Log(Shopping trips per UPC bought)	-0.12**	(0.00)			Yes	No	No
Log(Shopping trips per UPC bought)	-0.12**	(0.00)	0.10**	(0.04)	No	Yes	Yes
Log(Unique stores visited per $1K spent)	-0.23**	(0.01)			Yes	No	No
Log(Unique stores visited per $1K spent)	-0.22**	(0.01)	0.22**	(0.06)	No	Yes	Yes
Log(Unique stores visited per transaction)	-0.08**	(0.00)			Yes	No	No
Log(Unique stores visited per transaction)	-0.07**	(0.00)	0.30**	(0.06)	No	Yes	Yes
Log(Unique stores visited per brand bought)	-0.08**	(0.00)			Yes	No	No
Log(Unique stores visited per brand bought)	-0.08**	(0.00)	0.28**	(0.05)	No	Yes	Yes
Log(Unique stores visited per UPC bought)	-0.09**	(0.00)			Yes	No	No
Log(Unique stores visited per UPC bought)	-0.08**	(0.00)	0.29**	(0.05)	No	Yes	Yes
Instrumenting for household income:							
Log(Shopping trips per $1K spent)	-0.40**	(0.01)			Yes	No	No
Log(Shopping trips per $1K spent)	-0.40**	(0.01)	0.10**	(0.02)	No	Yes	Yes
Log(Shopping trips per transaction)	-0.29**	(0.01)			Yes	No	No
Log(Shopping trips per transaction)	-0.30**	(0.01)	0.19**	(0.03)	No	Yes	Yes
Log(Shopping trips per brand bought)	-0.26**	(0.01)			Yes	No	No
Log(Shopping trips per brand bought)	-0.27**	(0.01)	0.16**	(0.07)	No	Yes	Yes
Log(Shopping trips per UPC bought)	-0.28**	(0.01)			Yes	No	No
Log(Shopping trips per UPC bought)	-0.29**	(0.01)	0.18**	(0.06)	No	Yes	Yes
Log(Unique stores visited per $1K spent)	-0.35**	(0.01)			Yes	No	No
Log(Unique stores visited per $1K spent)	-0.34**	(0.01)	0.27**	(0.07)	No	Yes	Yes
Log(Unique stores visited per transaction)	-0.24**	(0.01)			Yes	No	No
Log(Unique stores visited per transaction)	-0.24**	(0.01)	0.37**	(0.06)	No	Yes	Yes
Log(Unique stores visited per brand bought)	-0.21**	(0.01)			Yes	No	No
Log(Unique stores visited per brand bought)	-0.20**	(0.01)	0.34**	(0.06)	No	Yes	Yes
Log(Unique stores visited per UPC bought)	-0.23**	(0.01)			Yes	No	No
Log(Unique stores visited per UPC bought)	-0.23**	(0.01)	0.35**	(0.06)	No	Yes	Yes

Note: Household income is calculated using the minimum of the range provided by Nielsen (e.g., a household with an income of $13K would be in the range $12–15K and would be assigned a value of $12K). Average county income is from the BEA Personal Income by County Area release. Grocery Estabs. are a count of NAICS 445 establishments from Census Business Patterns (includes grocery stores, supermarkets, liquor stores, and specialty food stores.) * indicates significance at 10%, ** at 5%.

Table B10: Comparison of markup distribution in data to model.

Percentile of markup distribution	\$20–\$25K Data	\$20–\$25K Model	\$50–\$60K Data	\$50–\$60K Model	\$100–\$125K Data	\$100–\$125K Model	Over \$200K Data	Over \$200K Model
10	0.83	1.13	0.84	1.13	0.88	1.13	0.93	1.13
25	1.01	1.15	1.02	1.15	1.06	1.16	1.10	1.17
50	1.21	1.20	1.22	1.20	1.27	1.23	1.33	1.26
75	1.45	1.32	1.46	1.33	1.52	1.41	1.60	1.51
90	1.76	1.58	1.77	1.61	1.85	1.78	1.94	2.00

Table B11: Calibration results varying number of UPC segments (K).

No. segments	Elasticity to others' incomes		Predicted Δ markup,
(K)	Markups paid	Search	1950–2018
1	0.056	0.062	14.1pp
5	0.038	0.036	10.8pp
10	0.036	0.035	10.6pp
20	0.035	0.034	10.5pp
50	0.035	0.034	10.4pp
100	0.035	0.034	10.4pp

Table B12: Predicted change in aggregate retail markup from 1950–2018, holding search effort by income constant.

Period	Predicted Δ in markup	Portion due to Δ Income level	Portion due to Δ Income dispersion
1950–2018	19.4pp	14.0pp	5.4pp
1950–1980	4.8pp	4.3pp	0.5pp
1980–2018	14.6pp	9.7pp	5.0pp

Table B13: Predicted change in aggregate retail markup from 1950–2018 under different values of reservation price R.

R	Predicted Δ in markup	Portion due to Δ Income level	Portion due to Δ Income dispersion
3.0	14.8pp	10.9pp	3.9pp
3.3 (Baseline)	14.1pp	10.4pp	3.7pp
4.0	13.1pp	9.7pp	3.4pp

Appendix C Proofs

C.1 Firms

Given $\{\bar{q}_n\}_{n=1}^{\infty}$, recall that $\bar{q}_1$ quotes are retrieved by households receiving only one quote, $2\bar{q}_2$ quotes are retrieved by households receiving two quotes, $3\bar{q}_3$ quotes are retrieved by households receiving three quotes, and so on. Hence, the demand facing a firm charging price $p \leq R$ is

$$\begin{aligned} D(p) &= \frac{1}{M}\left[\bar{q}_1 + 2\bar{q}_2\left(1 - F(p)\right) + 3\bar{q}_3\left(1 - F(p)\right)^2 + ...\right] \\ &= \frac{1}{M}\sum_{n=1}^{\infty} n\bar{q}_n\left(1 - F(p)\right)^{n-1}, \end{aligned}$$

and zero for any firm charging a price $p > R$. Accordingly, variable profits at any price $p \leq R$ are

$$\pi = \frac{1}{M}\left(p - mc\right)\sum_{n=1}^{\infty} n\bar{q}_n\left(1 - F(p)\right)^{n-1}.$$

Our equilibrium condition for the offer price distribution $F(p)$ is that all firms charging $p \in \text{supp}(F)$ make equal profits π, and any firm charging some $p \notin \text{supp}(F)$ will make profits strictly less than π. A firm charging the maximum price in the support of p (assuming $\bar{p} \leq R$) makes profits

$$\pi(\bar{p}) = \frac{1}{M}(\bar{p} - mc)\bar{q}_1.$$

As long as $\bar{q}_1 > 0$, profits of this firm are monotonically increasing in the price it charges in the region $p \leq R$, so it is clear that $\bar{p} = R$ as long as $\bar{q} > 1$. Hence, profits of all firms must be

$$\pi = \frac{1}{M}\left(R - mc\right)\bar{q}_1.$$

Accordingly, the distribution $F(p)$ is pinned down by the condition

$$\frac{R - mc}{p - mc}\bar{q}_1 = \sum_{n=1}^{\infty} n\bar{q}_n\left(1 - F(p)\right)^{n-1}.$$

Solving yields the expression for F (7) and for the minimum price $\underline{p}$ (8).

The aggregate markup (total sales over total costs) for firms in this economy can be represented as

$$\bar{\mu} = 1 + \frac{\int_{\underline{p}}^{R}\left(p - mc\right)D(p)\cdot MdF(p)}{\int_{\underline{p}}^{R} mcD(p)\cdot MdF(p)} = 1 + \int_{\underline{p}}^{R}\frac{\pi}{mc}\cdot MdF(p) = 1 + \left(\frac{R}{mc} - 1\right)\bar{q}_1. \tag{12}$$

Hence, given R/mc (i.e., the maximum markup in the support of F), the fraction of households getting only one price quote $\bar{q}_1$ is a sufficient statistic for the aggregate markup in the model. We can confirm this intuition by considering the two edge cases: when $\bar{q}_1 = 1$, no consumers search, resulting in the monopoly price equilibrium (Diamond 1971), and when $\bar{q}_1 = 0$, all consumers search, and we obtain competitive marginal cost pricing.

C.2 Household utility maximization

Given the distribution of prices offered by firms $F(p)$, the distribution of prices paid by household i is:

$$G_i(p) = \underbrace{q_{i,1}F(p)}_{\substack{\text{Receives one}\\ \text{price quote}}} + \underbrace{q_{i,2}(1-(1-F(p))^2)}_{\substack{\text{Receives two}\\ \text{price quotes}}} + \ldots = \sum_{n=1}^{\infty} q_{i,n}(1-(1-F(p))^n).$$

The average price paid by household i is then $p(s_i, F) = \int_{\underline{p}}^{R} p dG_i(p)$. Equivalently, the price function can be written as

$$p(s,F) = \sum_{n=1}^{\infty} q_n(s)\mathbb{E}[p|n],$$

where

$$\mathbb{E}\left[p|n\right] = \int_{\underline{p}}^{R} pn\left(1-F(p)\right)^{n-1} dF(p)$$

is the expected price paid having received n independent price quotes. For a non-degenerate distribution F, $\mathbb{E}[p|n]$ is strictly decreasing and convex in n (i.e., the returns to retrieving an additional price quote $(\mathbb{E}[p|n] - \mathbb{E}[p|n+1])$ are strictly positive and decreasing in n).

Using the above expression for $p(s,F)$ and the fact that $q_n = Q_n - Q_{n-1}$ (with the convention that $Q_{i,0} = 0$), we can rewrite the price function as

$$p(s,F) = \sum_{n=1}^{\infty} Q_n(s)(\mathbb{E}[p|n] - \mathbb{E}[p|n+1]).$$

Since F is fully pinned down by $\bar{q}$ in (7), I write the price function as $p(s,\bar{Q})$. It will be helpful to characterize the shape of the function $p(s,\bar{Q})$. First, note that the partial derivative of p with respect to search intensity s is

$$p_s = \sum_{n=1}^{\infty} \frac{dQ_n(s)}{ds}(\mathbb{E}[p|n] - \mathbb{E}[p|n+1]).$$

Under Assumption 1, $\frac{dQ_n(s)}{ds} \leq 0$ for all n and $\frac{dQ_1(s)}{ds} < 0$. Thus, when the distribution F is non-degenerate, $p_s < 0$.

Furthermore,

$$p_{ss} = \sum_{n=1}^{\infty} \frac{d^2Q_n(s)}{ds^2}(\mathbb{E}[p|n] - \mathbb{E}[p|n+1]).$$

Assumption 2 thus guarantees that $p_{ss} > 0$, i.e., the price function is decreasing and convex with respect to search intensity s.

Note that household first-order condition (5) is

$$\phi_i = -p_s(s, \bar{Q}).$$

Under Assumption 2, $-p_s$ is strictly decreasing, so we can invert this equation to write $s = s(\phi, \bar{Q})$. It will be helpful to note:

$$\frac{\partial s}{\partial \phi} = \frac{1}{-p_{ss}} < 0, \qquad \text{and} \qquad \frac{\partial^2 s}{\partial \phi^2} = \frac{p_{sss}}{-\left(p_{ss}\right)^3}.$$

C.3 Equilibrium stability

Suppose we are at an initial equilibrium where each household is choosing some search intensity $s_i(\phi, \bar{Q})$. Index this search intensity function by a parameter θ, such that s_i is strictly increasing in θ. In response to a perturbation in θ, the change in search intensity for any household i is

$$ds_i = \underbrace{\frac{\partial s_i}{\partial \theta} d\theta}_{\text{Direct effect}} + \underbrace{\frac{1}{-p_{ss}} \left(\nabla_{\bar{Q}} p_s\right) \cdot \left[\int_0^\infty \frac{dQ_n}{ds} \frac{\partial s(z)}{\partial \theta} d\Lambda(z)\right] d\theta}_{\text{Indirect effect through price distribution}},$$

where $\nabla_{\bar{Q}} p_s = \left[\frac{\partial p_s}{\partial \bar{Q}_1} \quad \ldots \quad \frac{\partial p_s}{\partial \bar{Q}_n} \quad \ldots\right]$ and (with some abuse of notation) $\left[\int_0^\infty \frac{dQ_n}{ds} \frac{\partial s(z)}{\partial \theta} d\Lambda(z)\right] = \left[\int_0^\infty \frac{dQ_1}{ds} \frac{\partial s(z)}{\partial \theta} d\Lambda(z) \quad \ldots \quad \int_0^\infty \frac{dQ_n}{ds} \frac{\partial s(z)}{\partial \theta} d\Lambda(z) \quad \ldots\right]'$. Thus, the change in the aggregate share of captive quotes received $\bar{Q}_1$ is

$$d\bar{Q}_1 = \underbrace{\left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{\partial s_i}{\partial \theta} d\Lambda(z)\right] d\theta}_{\text{Direct effect}} + \underbrace{\left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{1}{-p_{ss}} \nabla_{\bar{Q}} p_s(s, \bar{Q}) d\Lambda(z)\right] \cdot \left[\int_0^\infty \frac{dQ_n}{ds} \frac{\partial s(z)}{\partial \theta} d\Lambda(z)\right] d\theta}_{\text{Indirect effect through price distribution}}.$$

Define the aggregate returns to search Φ as

$$\Phi = \int_0^\infty -p_s(s(z), \bar{Q}) d\Lambda(z).$$

If differences in labor productivity across households are order $d\theta$, we can rewrite the indirect effect to a first order in terms of the change in the aggregate returns to search,

$$d\bar{Q}_1 = \underbrace{\left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{\partial s_i}{\partial \theta} d\Lambda(z)\right] d\theta}_{\text{Direct effect}} - \underbrace{\left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{1}{-p_{ss}} d\Lambda(z)\right] \frac{d\Phi}{d\bar{Q}_1} d\bar{Q}_1}_{\text{Indirect effect through price distribution}}.$$

Under Assumption 1 and Assumption 2, $p_{ss} > 0$ and $\frac{dQ_1}{ds} < 0$. If $d\Phi/d\bar{Q}_1$ is positive, then

$$\beta = \left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{1}{-p_{ss}} d\Lambda(z)\right] \frac{d\Phi}{d\bar{Q}_1} > 0.$$

So we re-arrange to get:

$$d\bar{Q}_1 = \frac{\left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{\partial s_i}{\partial \theta} d\Lambda(z)\right] d\theta}{1+\beta}. \tag{13}$$

Hence, we find that if $d\Phi/d\bar{Q}_1 > 0$, then (1) the change in $d\bar{Q}_1$ has the same sign of the direct effect $\left[\int_0^\infty \frac{dQ_1}{ds_i} \frac{\partial s_i}{\partial \theta} d\Lambda(z)\right] d\theta$, and (2) indirect effects dampen the overall change in $d\bar{Q}_1$ by a factor of $\frac{1}{1+\beta}$.

C.4 Comparative statics of aggregate markup to $\Lambda(z)$

Recall from (12) that the aggregate markup in the economy is

$$\bar{\mu} = 1 + \left(\frac{R}{mc} - 1\right)\bar{Q}_1.$$

From (13), to get the sign of $d\bar{Q}_1$ in response to a perturbation, we need to calculate only the sign of the direct effect of the perturbation on $\bar{Q}_1$.

Since $\bar{Q}_1 = \int_0^\infty Q_1(z) d\Lambda(z)$, a first-order stochastic shift in $\Lambda(z)$ increases $\bar{Q}_1$ if $Q_1(z)$ is increasing in z, and a mean-preserving spread in $\Lambda(z)$ increases $\bar{Q}_1$ if $Q_1(z)$ is increasing and convex in z. Hence, we need to show conditions under which $Q_1(z)$ is increasing and convex in z.

We can write,

$$\frac{dQ_1}{dz} = \frac{dQ_1}{ds}\frac{ds}{d\phi}\frac{d\phi}{dz} = \frac{dQ_1}{ds}\frac{1}{-p_{ss}}\frac{d\phi}{dz}.$$

Under Assumption 1(B), $\frac{dQ_1}{ds} < 0$, and under Assumption 2, $p_{ss} > 0$. Hence, $\frac{dQ_1}{dz} > 0$ if and only if $\frac{d\phi}{dz} > 0$.

The second derivative of Q_1 with respect to z is

$$\begin{aligned}
\frac{d^2Q_1}{dz^2} &= \frac{d^2Q_1}{ds^2}\left(\frac{ds}{d\phi}\frac{d\phi}{dz}\right)^2 + \frac{dQ_1}{ds}\frac{d^2s}{d\phi^2}\left(\frac{d\phi}{dz}\right)^2 + \frac{dQ_1}{ds}\frac{ds}{d\phi}\frac{d^2\phi}{dz^2} \\
&= \frac{d^2Q_1}{ds^2}\left(\frac{1}{-p_{ss}}\frac{d\phi}{dz}\right)^2 + \frac{dQ_1}{ds}\frac{p_{sss}}{-(p_{ss})^3}\left(\frac{d\phi}{dz}\right)^2 + \frac{dQ_1}{ds}\frac{1}{-p_{ss}}\frac{d^2\phi}{dz^2}.
\end{aligned}$$

Again using Assumptions 1 and 2, we can see that if $\frac{d^2\phi}{dz^2} > 0$, then a sufficient condition for $\frac{d^2Q_1}{dz^2} > 0$ is that

$$\frac{d^2Q_1}{ds^2} + \frac{dQ_1}{ds}\frac{p_{sss}}{-p_{ss}} \geq 0.$$

Rearranging yields,

$$\sum_{n=1}^{\infty}\left(\frac{d^2Q_{i,1}}{ds_i^2}\frac{d^2Q_{i,n}}{ds_i^2}-\frac{dQ_{i,1}}{ds_i}\frac{d^3Q_{i,n}}{ds_i^3}\right)\left[\mathbb{E}\left[p|n\right]-\mathbb{E}\left[p|n+1\right]\right]\geq 0.$$

which is exactly the condition guaranteed by Assumption 3. Thus, under Assumptions 1 and 2, $Q_1(z)$ is increasing in z if $\phi(z)$ is increasing in z, and hence a first-order stochastic shift in $\Lambda(z)$ increases $\bar{Q}_1$ and $\bar{\mu}$. Similarly, under Assumptions 1, 2, and 3, if $\phi(z)$ is increasing and convex in z, then $Q_1(z)$ is increasing and convex in z, and thus a mean-preserving spread in $\Lambda(z)$ increases $\bar{Q}_1$ and $\bar{\mu}$.

C.5 Application to two-quote and Poisson cases

I show that Assumptions 2 and 3 both hold under two common parameterizations of the search mapping function $\mathcal{S}$: a two-quote case and the Poisson case.

C.5.1 Application to two-quote case

Suppose that households always receive only one or two quotes, and that the probability of receiving two quotes is increasing in i's effort according to $q_{i,2} = 1 - \exp(-s_i)$. Thus

$$\begin{aligned} Q_{i,1} &= \exp\left(-s_i\right), \\ Q_{i,2} &= 1. \end{aligned}$$

I show that both Assumption 2 and Assumption 3 hold for this mapping $\mathcal{S}$.

Assumption 2 becomes:

$$\exp\left(-s_i\right)\left[\mathbb{E}\left[p|1\right]-\mathbb{E}\left[p|2\right]\right] > 0,$$

which is trivially true since $\exp\left(-s_i\right) > 0$ and since $\mathbb{E}\left[p|n\right]$ is strictly decreasing in n when F is non-degenerate.

Assumption 3 becomes:

$$\left(\left(-\exp\left(-s_i\right)\right)^2-\left(\exp\left(-s_i\right)\right)^2\right)\left[\mathbb{E}\left[p|1\right]-\mathbb{E}\left[p|2\right]\right] = 0 \geq 0.$$

So, we verify that the two-quote mapping satisfies both conditions.

The equilibrium condition in the two-quote case becomes

$$\left(\int_0^{\infty} q_1(z)d\Lambda(z)\right)\left(\mathbb{E}\left[p|1\right]-\mathbb{E}\left[p|2\right]\right) = \int_0^{\infty}\phi(z)d\Lambda(z). \tag{14}$$

With some algebra, we can show that the price distribution $F(p)$, the minimum price $\underline{p}$, the expected price from one and two searches $\mathbb{E}[p|1]$ and $\mathbb{E}[p|2]$, and the returns to search $\mathbb{E}[p|1] - \mathbb{E}[p|2]$ are given by:

$$
\begin{aligned}
F(p) &= 1 - \frac{1}{2}\frac{\bar{q}_1}{1-\bar{q}_1}\frac{R-p}{p-mc}, \\
\underline{p} &= mc + \frac{\bar{q}_1}{2-\bar{q}_1}\left(R-mc\right), \\
\mathbb{E}[p|1] &= mc + \frac{1}{2}\frac{\bar{q}_1}{1-\bar{q}_1}\left(R-mc\right)\log\left(\frac{2-\bar{q}_1}{\bar{q}_1}\right), \\
\mathbb{E}[p|2] &= mc + \frac{q_1}{1-\bar{q}_1}\left(R-mc\right) - \frac{1}{2}\left(\frac{q_1}{1-\bar{q}_1}\right)^2\left(R-mc\right)\log\left(\frac{2-\bar{q}_1}{\bar{q}_1}\right), \\
\mathbb{E}\left[p|1\right] - \mathbb{E}\left[p|2\right] &= \frac{\bar{q}_1}{1-\bar{q}_1}\left(R-mc\right)\left[\frac{1}{2}\frac{1}{1-\bar{q}_1}\log\left(\frac{2-\bar{q}_1}{\bar{q}_1}\right) - 1\right].
\end{aligned}
$$

So, we can rewrite (14) as

$$
\underbrace{\frac{\bar{q}_1^2}{1-\bar{q}_1}\left(R-mc\right)\left[\frac{1}{2}\frac{1}{1-\bar{q}_1}\log\left(\frac{2-\bar{q}_1}{\bar{q}_1}\right) - 1\right]}_{\text{Aggregate returns to search}} = \underbrace{\int_0^\infty \phi(z)d\Lambda(z)}_{\text{Aggregate cost of time}} .
$$

Figure C1a illustrates this aggregate equilibrium condition (using the exact functional form above). We can see that there is some threshold c such that (1) if the aggregate opportunity cost of time is greater than c, no dispersed-price equilibrium exists and the sole equilibrium is the monopoly-price equilibrium (all households choose $s_i = 0$); (2) if the aggregate opportunity cost of time is equal to c, there is exactly one value of $\bar{q}_1$ delivering a dispersed-price equilibrium, and (3) if the aggregate opportunity cost of time is less than c, there are two values of $\bar{q}_1$ with corresponding dispersed-price equilibria.

The arrows indicate how the equilibrium responds to a perturbation. We see that only the left-hand side equilibrium, where aggregate returns to search are increasing in $\bar{q}_1$, is a stable equilibrium. In this equilibrium, household decisions are strategic substitutes: an idiosyncratic increase in one household's search intensity decreases the returns to search and leads all other households to decrease search effort.

C.5.2 Application to Poisson distribution

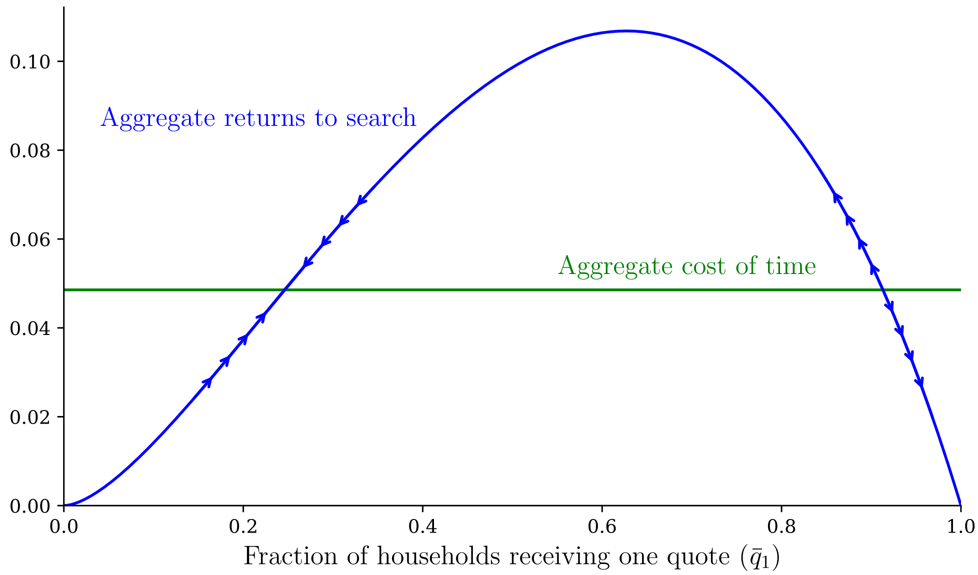
Under the Poisson distribution, the mapping from s_i to the probability mass function of price quotes is

$$
q_{i,n+1} = e^{-s_i}\frac{s_i^n}{n!}.
$$

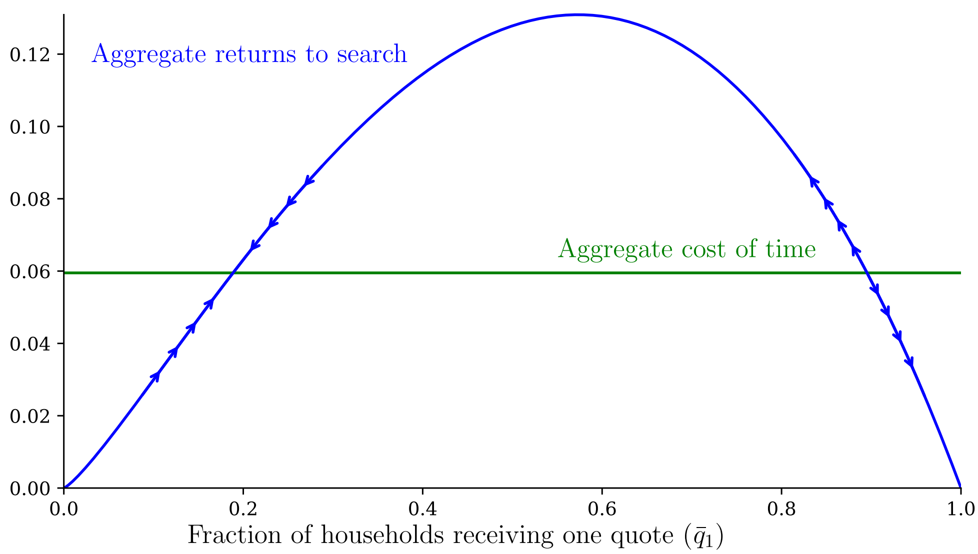
Note the index $n+1$, so that the support of the distribution starts from one. Accordingly,

$$
Q_{i,n+1} = \sum_{k=0}^{n} e^{-s_i}\frac{s_i^k}{k!}.
$$

Figure C1: Stable and unstable equilibria in two parameterizations of the nonsequential search equilibrium.



(a) Two-quote parameterization.



(b) Poisson parameterization.

For convenience in the below derivations, I drop the i subscripts. The first derivative with respect to s is

$$\begin{aligned}
\frac{dQ_{n+1}}{ds} &= -e^{-s}\left(1+\sum_{k=1}^{n}\frac{s^k}{k!}\right)+e^{-s}\left(\sum_{k=1}^{n}\frac{ks^{k-1}}{k!}\right) \\
&= -e^{-s}\left(1+\sum_{k=1}^{n}\frac{s^k}{k!}\right)+e^{-s}\left(1+\sum_{k=1}^{n-1}\frac{s^k}{k!}\right) \\
&= -e^{-s}\frac{s^n}{n!}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\frac{d^2Q_{n+1}}{ds^2} &= e^{-s}\frac{s^{n-1}}{n!}\left(s-n\right) && \text{for } n \geq 1, \\
\frac{d^3Q_{n+1}}{ds^3} &= -e^{-s}\frac{s^{n-2}}{n!}\left(\left(s-n\right)^2-n\right) && \text{for } n \geq 2.
\end{aligned}$$

For Q_1 and Q_2, we can explicitly write

$$\begin{aligned}
\frac{d^2Q_1}{ds^2} &= e^{-s} \\
\frac{d^3Q_1}{ds^3} &= -e^{-s}, \\
\frac{d^2Q_2}{ds^2} &= e^{-s}\left(s-1\right), \\
\frac{d^3Q_2}{ds^3} &= -e^{-s}\left(s-2\right).
\end{aligned}$$

Now we are ready to simplify Assumption 2:

$$\sum_{n=1}^{\infty}\frac{d^2Q_{i,n}}{ds_i^2}\left[\mathbb{E}\left[p|n\right]-\mathbb{E}\left[p|n+1\right]\right]>0$$

$$e^{-s}\left[\mathbb{E}\left[p|1\right]-\mathbb{E}\left[p|2\right]\right]+\sum_{n=1}^{\infty}\frac{d^2Q_{i,n+1}}{ds_i^2}\left[\mathbb{E}\left[p|n+1\right]-\mathbb{E}\left[p|n+2\right]\right]>0$$

With some algebra, this condition simplifies to:

$$\sum_{n=0}^{\infty}e^{-s}\frac{s^n}{n!}\left(\left[\mathbb{E}\left[p|n+1\right]-\mathbb{E}\left[p|n+2\right]\right]-\left[\mathbb{E}\left[p|n+2\right]-\mathbb{E}\left[p|n+3\right]\right]\right)>0$$

Since $\exp(-s)$ is strictly positive and $\mathbb{E}[p|n]$ is decreasing and convex in n, we verify this condition holds.

Now, for Assumption 3. Again, some algebra reveals:

$$\sum_{n=1}^{\infty}\left(\frac{d^2Q_{i,1}}{ds_i^2}\frac{d^2Q_{i,n}}{ds_i^2}-\frac{dQ_{i,1}}{ds_i}\frac{d^3Q_{i,n}}{ds_i^3}\right)[\mathbb{E}[p|n]-\mathbb{E}[p|n+1]]\geq 0.$$

$$\sum_{n=1}^{\infty}e^{-s}\left(\frac{d^2Q_{i,n}}{ds_i^2}+\frac{d^3Q_{i,n}}{ds_i^3}\right)[\mathbb{E}[p|n]-\mathbb{E}[p|n+1]]\geq 0.$$

$$[\mathbb{E}[p|2]-\mathbb{E}[p|3]]+\sum_{n=2}^{\infty}\left(\frac{s^{n-1}}{(n-1)!}-\frac{s^{n-2}}{(n-2)!}\right)[\mathbb{E}[p|n+1]-\mathbb{E}[p|n+2]]\geq 0.$$

$$\sum_{n=0}^{\infty}\frac{s^n}{n!}\left([\mathbb{E}[p|n+2]-\mathbb{E}[p|n+3]]-[\mathbb{E}[p|n+3]-\mathbb{E}[p|n+4]]\right)\geq 0.$$

Again, since $\mathbb{E}[p|n]$ is decreasing and convex in n, we verify this condition holds.

In the Poisson case, the aggregate equilibrium condition is simply

$$\sum_{n=1}^{\infty}\bar{q}_n\left(\mathbb{E}[p|n]-\mathbb{E}[p|n+1]\right)=\int_0^{\infty}\phi(z)d\Lambda(z).$$

Clearly, the left-hand side only depends on $\bar{q}$. Figure C1b illustrates this aggregate equilibrium condition (using the exact functional form above when individuals are identical). The implications are similar to those in the two-quote case.

C.6 Heterogeneous products extension

Households. Suppose there are K product segments. For each segment k, let v_{ik} denote household i's taste for segment k, given by

$$v_{ik}=\underbrace{\alpha_k(z_i)\delta_k}_{\substack{\text{Common to households}\\ \text{with income } z_i}}+\underbrace{\epsilon_{ik}}_{\text{Idiosyncratic}},$$

where ϵ_{ik} are i.i.d. draws from a Gumbel distribution. Given a households' tastes v_{ik}, household i's utility $u_i(\{c_{ik}\})$, where $\{c_{ik}\}$ is a vector of consumption by household i in each segment k, is

$$u_i(\{c_{ik}\})=\sum_k \mathbf{1}\{v_{ik}\geq v_{ik'},\ \forall k'\}\log(c_{ik}).$$

In words, household i only receives utility from the segment k for which it has the highest taste v_{ik}, and thus purchases exclusively from that segment when maximizing utility. A household's reservation price R_k is allowed to vary freely across segments, but, for a given segment k, is assumed to be identical across households.

Aggregating across households with income z, the fraction of households with income z that purchase from segment k is given by,

$$\text{Share}_k(z) = \frac{\exp(\alpha_k(z)\delta_k)}{\sum_{k'} \exp(\alpha_{k'}(z)\delta_{k'})}.$$

We can write the utility maximization problem of a household i that exclusively purchases from segment k as:

$$\max_{l_{ik}, s_{ik}} \log(c_{ik}) \qquad \text{s.t.} \qquad \begin{cases} t_i(c_{ik}, s_{ik}) + l_{ik} = 1, & \text{(Time constraint)} \\ p_{ik}c_{ik} = z_i l_{ik}, & \text{(Budget constraint)} \end{cases}$$

where $t_i(c_{ik}, s_{ik})$ is the time it takes household i to shop for c_{ik} units with search intensity s_{ik}, l_{ik} is time spent working with labor productivity z_i, and p_{ik} is the average price paid by i per unit of consumption in segment k. Again, the budget constraint anticipates free entry, so that all household earnings come from labor market work.

Thus, each household chooses search intensity s_{ik} according to the first-order condition,

$$-\frac{\partial p(s_{ik}, F_k)}{\partial s_{ik}} = \phi_i,$$

where the opportunity cost of search effort $\phi_i = z_i/a_i$ is defined in the same way as in the main text. Notice that the household's decision on search intensity now depends on which segment it purchases from, because the returns to search depend on the distribution of prices posted in that segment, which are given by F_k.

Given a distribution of labor productivity $H(z)$, the distribution of buyers' incomes in each segment is given by

$$d\Lambda_k(z) = \frac{c_{ik}}{C_k}\text{Share}_k(z)dH(z),$$

where aggregate consumption in the segment C_k is $C_k = \int_0^\infty c_{ik}\text{Share}_k(z)dH(z)$. Since each segment may have a different distribution of buyers' incomes $\Lambda_k(z)$, households with the same income may choose different search intensities in each segment, and the aggregate search behavior within each segment will be different. In particular, aggregate search behavior in segment k is

$$\bar{q}_{nk} = \int_0^\infty q_k(z)d\Lambda_k(z), \qquad \text{for all } n.$$

Firms. Firms each operate in a single segment, and pay an entry cost f_k in units of labor to enter segment k. Free entry and exit determines the mass of firms in each segment, M_k, so that in equilibrium variable profits earned in segment k satisfy

$$\pi_k(p) - f_k \leq 0, \qquad \text{for all } p$$

In each segment, firms produce with a constant-returns production technology with productivity A_k, so that the marginal cost of producing one unit of the segment output is $1/A_k$. The distribution of prices in segment k follows closely from the model presented in the main text: Given $\{\bar{q}_{nk}\}_{n=1}^{\infty}$ with $\bar{q}_{1k} \in (0,1)$, the unique equilibrium price distribution $F_k(p)$ is

$$F_k(p) = \begin{cases} 0 & \text{if } p < \underline{p}_k \\ 1 - \Psi_k \left[\left(\frac{R_k - 1/A_k}{p - 1/A_k} \right) \bar{q}_1 \right] & \text{if } \underline{p}_k \leq p \leq R_k \\ 1 & \text{if } p > R_k \end{cases}$$

where the lowest price $\underline{p}_k$ in the support of F_k is

$$\underline{p}_k = 1/A_k + \frac{\bar{q}_{1k}}{\sum_{n=1}^{\infty} n\bar{q}_{nk}} (R_k - 1/A_k),$$

and $\Psi_k(\cdot)$ is the inverse of the strictly increasing, C^{∞} function $y(x) = \sum_{n=1}^{\infty} n\bar{q}_{nk} x^{n-1}$.

As in the single-product model, we can use the profits of the firm with price R_k in each segment to show that the aggregate markup in segment k is

$$\bar{\mu}_k = 1 + (R_k A_k - 1)\,\bar{q}_{1k}.$$

Equilibrium. An equilibrium is defined in a way analogous to the single-product model: each household purchases exclusively from the segment k for which they have highest per-purchase utility and chooses search intensity to maximize utility, each firm in segment k maximizes profits by choosing a price $p \in \text{supp}(F_k)$, the number of firms in each segment M_k is such that variable profits $\pi_k(p) = f_k$ for any $p \in \text{supp}(F_k)$, and all markets clear.

Appendix D Changes in Search Behavior over Time

In this appendix, I use three measures—time spent searching per purchase, total time allocated to search, and price dispersion—to assess how the market's predictions compare to the data. For all three measures, the model appears to match the data on search behavior and price dispersion over time reasonably well.

Comparison of search time per good consumed. Table D1 reports the change in search productivity and time spent shopping per purchase for various quantiles of the income distribution from 1950 to 2018. For all quantiles of the income distribution, the model predicts a large increase in search productivity and a consequent decline in time spent shopping per purchase. For example, search productivity of a household with median earnings more than doubles from 1950 to 2018 and shopping time per purchase declines over 50%. For individuals at the 90th percentile of earnings, search productivity increases 110% and time per purchase declines by nearly two-thirds. (The growth is more pronounced at the high end of the income distribution because income, and by implication labor productivity, has also grown fastest for the high end.)

How do these calculations within the model compare with empirical evidence on search behavior over time? I take the following specification to Nielsen Homescan data from 2004–2019:

$$\text{SearchPerPurchase}_{it} = \delta_t + \alpha_i + \gamma' X_{it},$$

where δ_t are year fixed effects, α_i are household fixed effects, and X_{it} includes additional controls for household age and county (to control for households' moving across counties or for changes over the lifecycle). Assuming that households stay at the same quantile of the income distribution over time, the fixed effects δ_t will reflect secular changes in shopping time per purchase as in the final column of Table D1.

Figure D1 plots the fixed effects on each year (relative to 2004) for four proxies of shopping time per purchase. Two measures indicate a reduction in shopping time per purchase on the order of 10–15 percent. For comparison, the decrease in shopping time per purchase in the model over this period in the model is between 5–20 percent (depending on quantile of the income distribution). Hence, these measures align well with the predictions of the model. However, comparing across measures, the evidence of a systematic decline in shopping time per purchase in the data is not conclusive.[55]

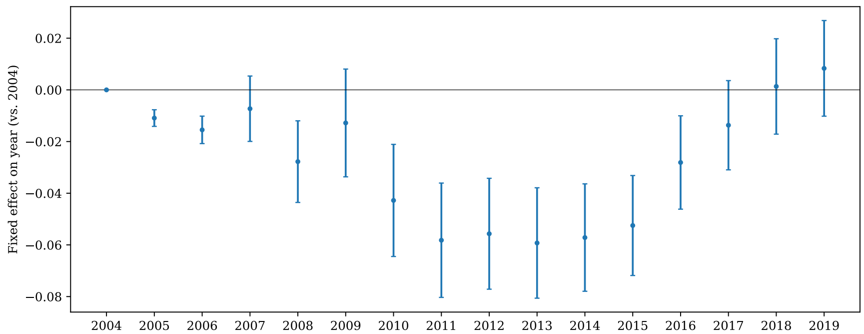
Comparison of total time allocated to obtaining goods and services. In the model, two opposing forces affect the total amount of time a household allocates to shopping. First, when opportunity

[55] If time spent shopping per purchase declines less than predicted by the model, this could suggest that search productivity has not fallen as much in reality as is assumed in the model. However, it could also result from search intensity not falling as much as the model predicts. Hence, it is challenging to interpret whether this pattern in the data suggests that search productivity improvements in the model are aggressive—which would suggest that the model's predictions are conservative—or that the decline in search intensity predicted by the model overshoots the data.

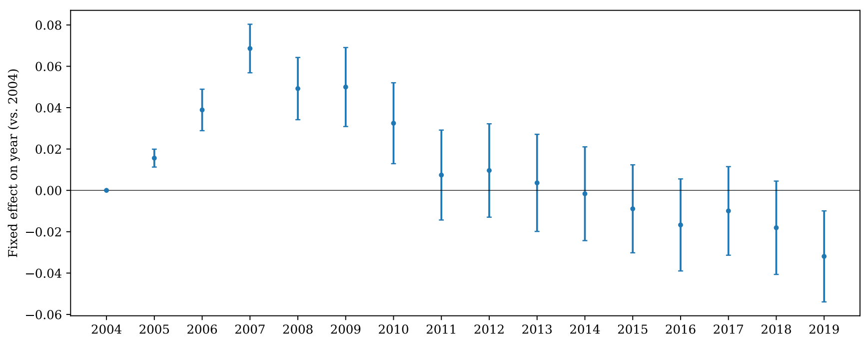
Table D1: Change in shopping productivity and shopping time by percentile of income distribution relative to 1950.

		Change relative to 1950 (as a percent)		
Percentile	Year	Post-tax income (z_i)	Search productivity (a_i)	Shopping time per purchase
25	1962	46.3	54.0	-31.2
	1970	72.3	72.7	-40.2
	1980	73.7	73.5	-40.1
	1990	71.3	72.1	-37.7
	2000	98.6	87.8	-44.0
	2010	102.1	89.8	-44.7
	2018	107.7	95.2	-45.7
50	1962	44.7	28.9	-26.8
	1970	68.9	52.9	-37.5
	1980	77.1	63.7	-40.4
	1990	96.2	88.3	-45.2
	2000	130.7	104.5	-51.0
	2010	146.1	120.5	-54.0
	2018	162.2	136.4	-56.4
75	1962	44.8	52.7	-30.1
	1970	72.2	73.4	-40.5
	1980	83.4	85.7	-43.8
	1990	111.0	104.6	-49.1
	2000	156.1	125.4	-55.8
	2010	185.6	134.6	-59.6
	2018	213.8	141.5	-62.4
90	1962	46.9	34.1	-28.4
	1970	75.2	46.8	-38.7
	1980	89.4	51.8	-42.2
	1990	125.6	60.5	-48.8
	2000	192.3	78.4	-59.0
	2010	220.2	91.5	-62.3
	2018	261.4	111.1	-66.4

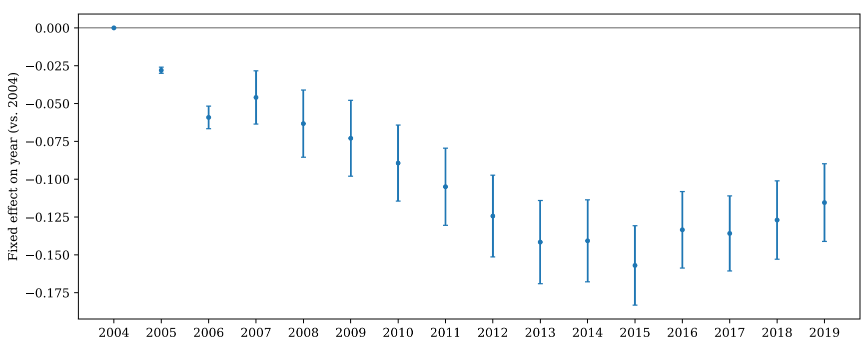
Figure D1: Proxies for search time over years 2004–2019.



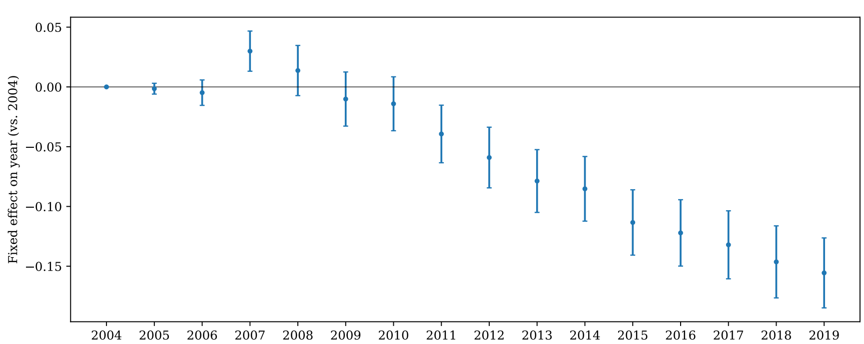
(a) Log(Shopping trips per $1,000 2007 USD spent).



(b) Log(Unique retailers visited per $1,000 2007 USD spent).



(c) Log(Shopping trips per unique brand purchased).



(d) Log(Unique retailers visited per unique brand purchased).

cost of search effort is increasing in income, high-income households spend less time on search per purchase. However, high-income households also shop for a larger consumption bundle. Thus, the total amount of time allocated to shopping need not be monotonic in income. In the model, this can be seen by noting that the ratio of time spent working to time spent on market labor is equal to the elasticity of the price function with respect to search intensity,

$$\frac{t_i}{l_i} = \frac{1}{l_i}\frac{c_i}{a_i}s_i = \frac{z_i}{a_i}\frac{s_i}{p_i} = -\frac{\partial \log p}{\partial \log s}.$$

Since this elasticity is non-monotonic in search intensity (as shown in Figure B14), high-income households may spend more or less time shopping than low-income households. Hence, it is challenging to ascertain whether the model is aggressive or conservative relative to the data using measures of total amount of time allocated to shopping.

With this caveat in mind, I compare the ratio of time spent shopping to time spent working in the calibrated model to data reported by Aguiar and Hurst (2007b) from 1965 to 2003. From time use surveys over several decades, Aguiar and Hurst (2007b) find that the hours per week allocated to both obtaining goods and services and to total market work are both declining. Accordingly, the ratio of time spent obtaining goods and services to time spent on total market work is roughly stable, fluctuating between 0.15 and 0.18 from 1965 to 2003.[56] For comparison, the ratio of average time spent shopping to time spent working in the calibrated model is 0.144 in 1950, and rises slightly to 0.152 by 2018. My assessment is that the ratio of time spent shopping to time spent on market work in the model is well-matched to its empirical counterpart, and that the change in this ratio in the model, as in the data, is negligible.

Comparison of price dispersion. Finally, I check how the evolution of price dispersion in the model compares to available empirical evidence. One might be concerned that the decline in aggregate search intensity in the model generates a significant increase in price dispersion, at odds with the empirical evidence. However, Table D2 shows that changes in price dispersion predicted by the model are small. In particular, the model predicts less than a 2 percent change in the standard deviation of posted prices from 1950–2018, and the change is not monotonic over time.

This evidence accords with a literature that has measured price dispersion over time and across media (e.g., in-store versus online). As noted by Menzio (2021), despite innovations that have presumably increased consumers' searxh productivities, price dispersion as measured by Pratt et al. (1979) in the late 1970s, Lach (2002) in 1993, and Kaplan and Menzio (2015) in 2010 is relatively constant. In fact, Scholten and Smith (2002) finds that price dispersion in e-commerce markets in 2000 was slightly higher than price dispersion among brick-and-mortar retailers in 1976. Similarly, Brynjolfsson and Smith (2000) and Baye et al. (2004) find that price dispersion in online retail is comparable to offline, despite presumably lower frictions in search online.

[56] A regression of the ratio of shopping time to market work time on year yields a noisy zero. I also conduct a similar analysis using data from the American Time Use Survey from 2003 to 2019; there is also no discernable trend in the the ratio of shopping time to market work time in ATUS data over this period.

Table D2: Predicted change in the dispersion of posted prices from 1950–2018.

Year	Δ Standard deviation of posted prices, relative to 2010
1950	-1.86%
1962	-1.03%
1970	-1.09%
1980	-0.89%
1990	-0.16%
2000	0.98%
2010	0.00%
2018	-0.08%

Appendix E Sequential Search Model

As an alternative to nonsequential search, in this appendix I describe a model where households search for products sequentially. The model parallels the labor market sequential search model by Burdett and Mortensen (1998). Again, the innovation in the model is that heterogeneous households endogenously choose search intensities.

E.1 Households

There is a unit measure of households indexed by type $i \in [0, \infty)$. Types are distributed in the population according to the density $dH(i)$, where $H(i)$ is the share of households with type less than or equal to i. Households search for an identical good sold by a measure of M firms. All households are risk-neutral and discount future utility at rate r. As in canonical models of search frictions, households know the distribution of prices offered by firms, $F(p)$, but do not know which retailer sells at which price. Denote $\underline{p}$ and $\bar{p}$ the infimum and supremum of the support of F.

At any moment, a household is either "matched" to a retailer or unmatched. Matching can be thought of as a consumption habit—for instance, a household may be used to buying milk from a certain retail outlet. At an arrival rate λ_i, household i receives information about the price of the good sold by another retailer. (As we will see later, the arrival rate of new price quotes is a result of i's endogenous choice of search intensity.) Since households search randomly across retailers, this new price quote is assumed to be a random draw from $F(p)$. Households have no recall of previous price quotes and may only switch to buying from the new retailer at the time when the quote arrives. Matches between households and retailers are destroyed at an exogenous positive rate δ (which can be interpreted as discontinued products, price changes, or store closures).

At time t, household i's flow utility is

$$u_{i,t} = \begin{cases} z_i(T - t_i) + R - p_{i,t} & \text{if } i \text{ purchases good at price } p_{i,t} \text{ in period } t \\ z_i(T - t_i) & \text{otherwise} \end{cases}$$

where t_i is the time i spends shopping, $T - t_i$ is the time i spends working with wage equal to i's labor productivity z_i, and R is the value of the good (which I assume is identical across households).

Given this setup, the expected discounted lifetime utility of household i when unmatched, $V_{i,0}$, and when matched to a retailer offering the good at price p, $V_{i,1}(p)$, satisfy

$$rV_{i,0} = z_i (T - t_i) + \lambda_i \left[\int_{\underline{p}}^{\bar{p}} \max \{V_{i,1}(p) - V_{i,0}, 0\} \, dF(p) \right],$$

$$rV_{i,1}(p) = z_i (T - t_i) + R - p + \lambda_i \left[\int_{\underline{p}}^{\bar{p}} \max \{V_{i,1}(x) - V_{i,1}(p), 0\} \, dF(x) \right] + \delta \left[V_{i,0} - V_{i,1}(p) \right].$$

It is straightforward to show that $V_{i,1}(p) \geq V_{i,0}$ only if $p \leq R$. For this reason, R is the maximum price at which any household is willing to buy the good. As I will show below, no firm chooses to set a price above R in equilibrium, so we can simplify all subsequent expressions using $F(R) = 1$.

I will now proceed in two steps. First, I show that if there is some p in the support of F such that $\underline{p} < p < \bar{p}$ (the distribution of prices F has more than two unique points in its support), then the steady-state distribution of prices paid by a household i first-order stochastically dominate the prices paid by another household j if and only if $\lambda_i < \lambda_j$. Second, I endogenize each household's search intensity decision and derive sufficient conditions under which search intensity is decreasing with i.

Denote the steady-state distribution of prices paid by household i by $G_i(p)$ and the fraction of households of type i that are unmatched to a retailer at any moment by ω_i. In steady state, the flows of households of type i to a retailer with price greater than or equal to p must equal the out-flows of households of type i matched to a retailer with price greater than or equal to p:

$$\lambda_i \left[F(R) - F(p)\right] \omega_i dH(i) = \left[\delta + \lambda_i F(p)\right](1 - \omega_i)\left[1 - G_i(p)\right] dH(i).$$

Setting $p = \underline{p}$ allows us to solve for the unmatched fraction of households of type i, ω_i, and thus the distribution of prices paid by household i:

$$G_i(p) = \frac{(\delta + \lambda_i) F(p)}{\delta + \lambda_i F(p)}.$$

Consider two households i and j where $\lambda_i < \lambda_j$. The distribution of prices paid by household i first order stochastically dominates that paid by household j if $G_i(p) \leq G_j(p)$ for all p and $G_i(p) < G_j(p)$ for some p. Since

$$G_j(p) = G_i(p) + \frac{\delta F(p)(1 - F(p))}{\left[\delta + \lambda_i F(p)\right]\left[\delta + \lambda_j F(p)\right]}\left(\lambda_j - \lambda_i\right), \tag{15}$$

it is clear that $G_j(p) \leq G_i(p)$ for all p and that the inequality will hold strictly for any p where $F(p)$ is not zero or one. Equation (15) conveys the main intuition for cross-sectional differences in prices paid across households due to differences in search intensity: by increasing search intensity λ_i, a household shifts the steady-state distribution of prices it pays to the left.

Households choose to exert search effort to maximize expected discounted lifetime utility.

$$\lambda_i = \arg\max_{\lambda} \mathbb{E}\left[\int_0^{\infty} \exp(-rt)\ u_{i,t}\ dt\right].$$

I assume that the time required to achieve search intensity λ_i is given by the concave function

$$\lambda_i = 1 - \exp(-a_i t_i),$$

where a_i is search productivity that can vary across households. The first-order condition for λ_i

equates the opportunity cost of increasing search effort to the expected benefit from doing so:

$$\phi_i = \delta\,(1-\lambda_i)\left[\frac{1}{(\delta+\lambda_i)^2}R - \int_{\underline{p}}^{R} \frac{\delta - \lambda_i F(p)}{[\delta+\lambda_i F(p)]^3} p\,dF(p)\right], \tag{16}$$

where $\phi_i = z_i/a_i$ is the opportunity cost of search effort.

With some abuse of notation, I assume that there is a one-to-one mapping between household type i and labor productivity z_i. Hence, we can write the above equation as:

$$\phi(z) = \delta\,(1-\lambda(z))\left[\frac{1}{(\delta+\lambda(z))^2}R - \int_{\underline{p}}^{R} \frac{\delta - \lambda(z) F(p)}{[\delta+\lambda(z) F(p)]^3} p\,dF(p)\right].$$

Taking the comparative static with respect to z yields Proposition 3.

Proposition 3. *Search intensity $\lambda(z)$ is decreasing in z if $\phi(z)$ is increasing in z.*

Now that we know how the prices paid by a household depend on its search intensity, and how that search intensity is determined by a household's labor and search productivity, we can move on to the firm problem.

E.2 Firms

A measure M of ex ante identical firms with marginal cost mc set prices to maximize profits. The demand that a firm faces depends on its price and on the distribution of prices charged by other firms. In particular, the demand from household with labor productivity z at a price $p \le R$ is

$$\begin{aligned} D_z(p) &= \frac{1}{M}\frac{G_i(p^+) - G_i(p)}{F(p^+) - F(p)}\,(1-\omega(z))\,dH(z) \\ &= \frac{\delta}{M}\frac{\lambda(z)}{[\delta+\lambda(z)F(p^+)]\,[\delta+\lambda(z)F(p)]}dH(z), \end{aligned}$$

where p^+ is a price marginally greater than p. Demand at a price $p > R$ is zero since R is the reservation price for all households.

Aggregating across households, we find that a firm charging price $p \le R$ has profits

$$\pi(p) = (p - mc)\,\frac{\delta}{M}\int_0^{\infty} \frac{\lambda(z)}{[\delta+\lambda(z)F(p^+)]\,[\delta+\lambda(z)F(p)]}dH(z).$$

The price distribution F is an equilibrium price distribution if firms charging p in the support of F make identical profits π, but any firm charging $p \notin \mathrm{supp}(F)$ makes profits strictly less than π. As long as $R > mc$, $\pi(R) > 0$ and so the maximum price in the support of F must be less than or equal to R. As such, we can limit our focus to distributions F where $\bar{p} \le R$.

We can also rule out non-continuous distributions for F. The intuition is the same as in the Burdett and Mortensen (1998) model with identical workers: if F has a mass point at some $\hat{p}$, then a firm offering a price slightly lower than $\hat{p}$ will have significantly higher demand and only a marginal loss in profits per item sold.

Consider the profits of a firm charging that maximum price $\bar{p}$:

$$\pi(\bar{p}) = (\bar{p} - mc)\,\frac{\delta}{M}\int_0^\infty \frac{\lambda(z)}{(\delta + \lambda(z))^2}dH(z). \tag{17}$$

Clearly, $\pi(\bar{p})$ is strictly increasing in $\bar{p}$, so the maximum price will exactly equal the households' reservation price R. Since $\pi(p) = \pi(R)$ for all $p \in \text{supp}(F)$, we can pin down the minimum price in F and the overall shape of F:

$$\underline{p} = mc + \delta^2 (R - mc)\,\frac{\int_0^\infty \frac{\lambda(z)}{(\delta+\lambda(z))^2}dH(z)}{\int_0^\infty \lambda(z)dH(z)} \qquad \text{and} \qquad \frac{\int_0^\infty \frac{\lambda(z)}{(\delta+\lambda(z))^2}dH(z)}{\int_0^\infty \frac{\lambda(z)}{\left[\delta+\lambda(z)F(p)\right]^2}dH(z)} = \frac{p - mc}{R - mc}. \tag{18}$$

E.3 Equilibrium

Given R, mc, and household distribution $H(z)$, an equilibrium is a tuple $\left(\{\lambda(z)\}_{z=0}^{\infty}, F, \pi, M\right)$ where search intensity $\lambda(z)$ maximizes expected discounted lifetime utility given F for all z, all firms choosing a price $p \in \text{supp}(F)$ have profit π given the prices charged by other firms F, any price $p \notin \text{supp}(F)$ results in profits that are strictly less than π, and M is such that the zero profit condition holds.

Equivalently, $\lambda(z)$ satisfies (16) for all z; π satisfies (17) with $\bar{p} = R$; M is such that $\pi = f_e$; and $F(p)$ is given by (18) for all $p \in [\underline{p}, R]$, is zero for $p < \underline{p}$ given in (18), and is one for $p > R$.

In equilibrium, the aggregate markup is

$$\bar{\mu} = 1 + \delta\left(\frac{R}{mc} - 1\right)\frac{\int_0^\infty \frac{\lambda_i}{(\delta+\lambda_i)^2}dH(i)}{\int_0^\infty \frac{\lambda_i}{\delta+\lambda_i}dH(i)}.$$

E.4 Calibration

I calibrate the model using an analogous two-step procedure to the one described in the main text. Importantly, I find that the results in the sequential search model are more sensitive to the valuation R. So, I choose R by minimizing the distance between the 50th and 75th percentiles of the markup distribution for each household type and their empirical counterparts. I find $R = 2.16$, which is lower than the value used in the nonsequential search calibration. I choose $\delta = 0.10$; the results are not sensitive to the choice of δ.[57]

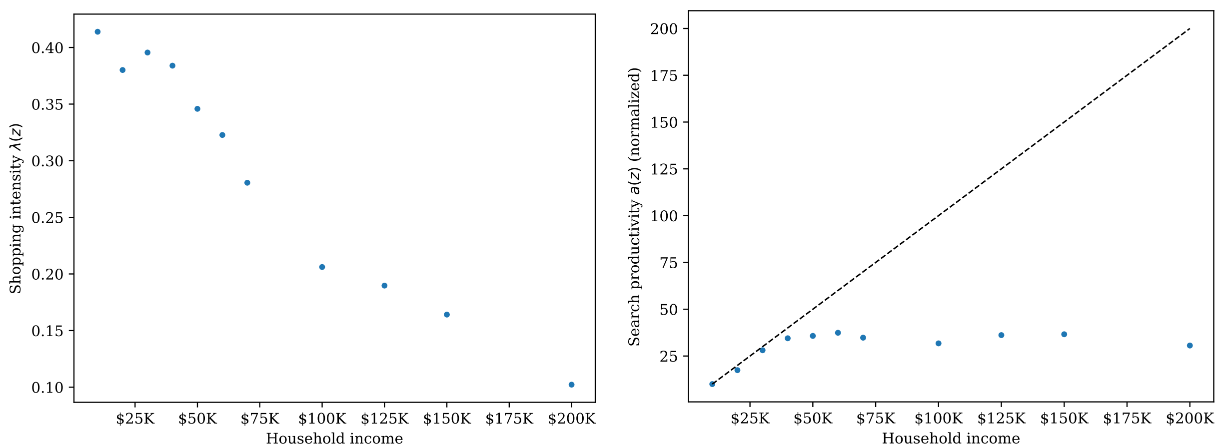
I reproduce results from the main text in this calibration of the sequential search model: Figure E1 shows the calibrated search intensities ($\lambda(z)$) and shopping productivities ($a(z)$); Figure E2 shows the model's prediction of the change in markups over time, holding all factors other than the

[57] As can be seen in the model derivation, $\lambda(z)/\delta$ drives the behavior of the price distribution and the aggregate markup. So, the calibration chooses values of $\lambda(z)/\delta$ that fit the data, and the choice of δ simply normalizes the results.

Table E1: Predicted change in aggregate retail markup from 1950–2018 under sequential search model.

Period	Predicted Δ in markup	Due to Δ Income level	Due to Δ Income dispersion	Due to Within-firm changes	Due to Cross-firm reallocations
1950–2018	12.8pp	9.2pp	3.6pp	8.0pp	4.8pp
1950–1980	2.9pp	2.6pp	0.3pp	1.7pp	1.2pp
1980–2018	9.8pp	6.6pp	3.3pp	6.3pp	3.5pp

Figure E1: Calibrated shopping intensity $\lambda(z)$ and search productivity $a(z)$.



(a) Search intensity $\lambda(z)$.

(b) Search productivity $a(z)$.

income distribution constant; and Figure E3 reproduces additional facts about the decomposition of the rise in markups into within-firm changes and cross-firm reallocations. The results are qualitatively similar to the results of the nonsequential search calibration presented in the main text.

Figure E2: Predicted aggregate retail markup under income distributions from 1950–2018.

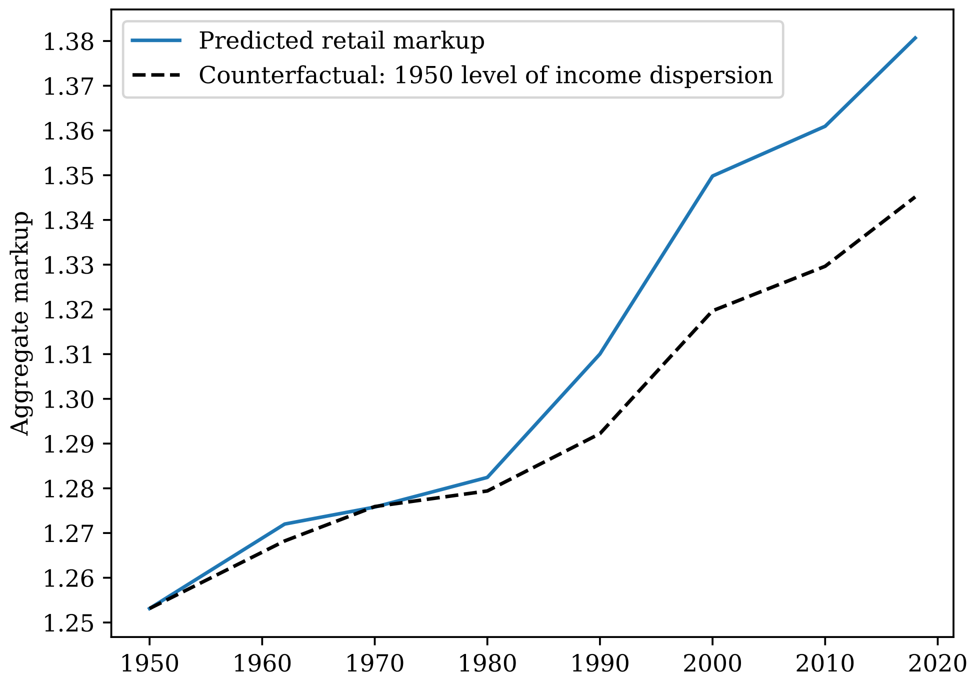
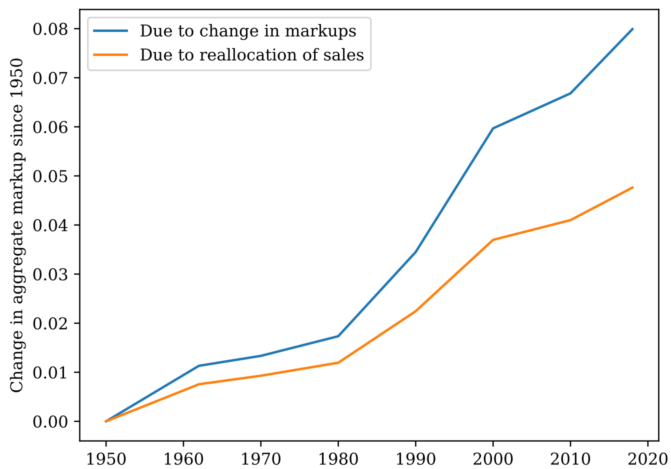
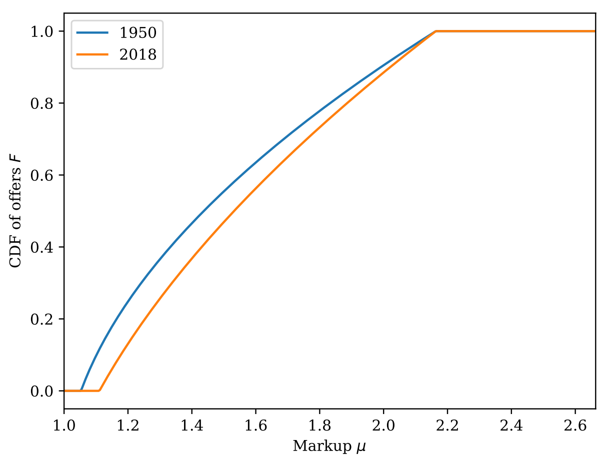


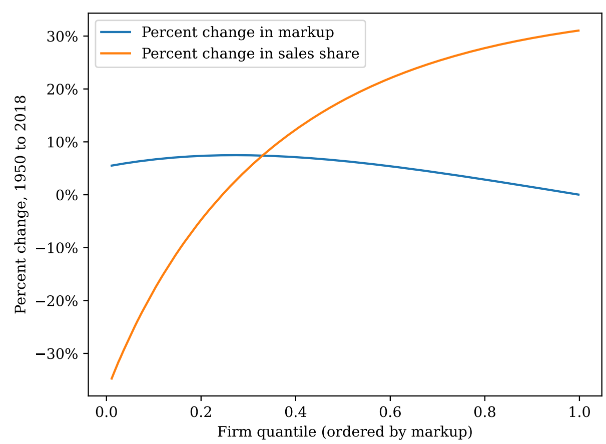
Figure E3: Predictions of search model under income distribution from 1950–2018.



(a) Decomposition of change in agg. markup.



(b) Offer distribution F in 1950 and 2018.



(c) Percent change in markups and sales shares by quantile from 1950 and 2018.

Appendix F Differences in Elasticities of Substitution and Taste for Quality

In the model presented in the main text, differences in price sensitivity across income groups arise endogenously from search effort. An alternative micro-foundation is that differences in price sensitivity arise due to differences in preferences across households. As discussed by Handbury (2021), differences in taste for quality and differences in the elasticity of demand with respect to price are two common forms of non-homotheticity explored in the urban and international literatures. Hence, in this appendix, I present a model in which households have preferences that differ in terms of taste for quality and elasticities of substitution. A calibration of this model suggests an increase in the aggregate retail markup from 1950–2018 of 15.7pp. This is larger than the results presented in the main text because this version of the model does not feature a dampening feedback loop from search behavior.

Model setup. Households consume varieties indexed by type $k \in [0,1]$. Household i's utility from consuming c_{ik} units of each variety of type k is given by

$$u_i = \left[\int_0^1 (q_{ik} c_{ik})^{\frac{\sigma_i - 1}{\sigma_i}} \, dF(k) \right]^{\frac{\sigma_i}{\sigma_i - 1}}, \tag{19}$$

where $F(k)$ is the density of varieties available with type k, q_{ik} is a household-specific quality shifter for type k, and $\sigma_i > 1$ is household i's elasticity of substitution. I assume that all households with income z_i share the same quality shifters and elasticities of substitution, so that we can write

$$q_{ik} = q_k(z_i), \qquad \sigma_i = \sigma(z_i).$$

The preferences in (19) are based on non-homothetic preferences used by Handbury (2021) and Faber and Fally (2022), though there is a difference in micro-foundation. For ease of exposition, I have assumed that differences in taste for quality and elasticities of substitution across households are because households at different income levels have different types and thus different preferences altogether. Handbury (2021) instead assumes all households share common preferences, but that households' utilities for goods may depend on their level of non-grocery expenditures. (Complementarities between non-grocery expenditures and quality, for example, generate non-homotheticities in taste for quality.) While the micro-foundations differ, the type-specific preferences in (19) allow taste for quality and elasticities of substitution to vary with income in a way that parallels the non-homothetic preferences in Faber and Fally (2022) and Handbury (2021).

Households supply labor inelastically, receiving z_i in wages, and choose consumption of each type k to solve the maximization problem,

$$\max_{\{c_{ik}\}} u_i \qquad \text{s.t.} \qquad \int_0^1 p_k c_{ik} dF(k) = z_i.$$

Solving the maximization problem yields household i's demand for a variety of type k,

$$q_{ik} c_{ik} = \left(\frac{p_k}{q_{ik}}\right)^{-\sigma_i} \frac{z_i}{\int_0^1 \left(\frac{p_k}{q_{ik}}\right)^{1-\sigma_i} dk}.$$

That is, household i's elasticity of demand for a variety of type k is σ_i, and consumption of variety k increases with the household-specific quality shifter q_{ik} and with income z_i.

Each firm supplies a single variety and sets price to maximize profits, taking as given all other prices. Firms produce with a constant-returns technology with a per-unit cost of $1/A_k$. Thus, firm k maximizes

$$\max_{p_k} (p_k - \frac{1}{A_k}) C_k,$$

where total demand C_k is

$$C_k = \int_{\mathcal{I}} c_{ik} di,$$

and $\mathcal{I}$ is the set of all household types. Using the household demand curves above, the profit-maximizing price for firm k satisfies

$$p_k = \mu_k \frac{1}{A_k} = \frac{1}{1 - 1/\sigma_k} \frac{1}{A_k}, \qquad \text{where} \qquad \sigma_k = \frac{\int_{\mathcal{I}} \sigma_i p_k c_{ik} di}{\int_{\mathcal{I}} p_k c_{ik} di}.$$

Intuitively, a firm's price is equal to its marginal cost times a markup. The profit-maximizing markup is given by the usual Lerner formula, using the aggregate elasticity of demand σ_k facing the firm, which is simply the expenditure share-weighted average of k's consumers' elasticities of demand. (Since the aggregate elasticity of demand σ_k depends on the firm's price p_k, the profit-maximizing price p_k solves a fixed point.)

There is free entry. Firms pay a fixed entry cost f_e in units of labor and subsequently realize their type k. The zero profit condition means that expected post-entry profits equal the fixed cost,

$$\int_0^1 (p_k - \frac{1}{A_k}) C_k dF(k) = f_e.$$

In equilibrium, households choose consumption $\{c_{ik}\}$ to maximize utility, firms choose prices to maximize profits, the zero profit condition holds, and resource constraints are satisfied.

For calibrating the model, it will be useful to denote the share of income that household i spends on variety k as χ_{ik},

$$\chi_{ik} = \frac{p_k c_{ik}}{\int_0^1 p_k c_{ik} dF(k)}.$$

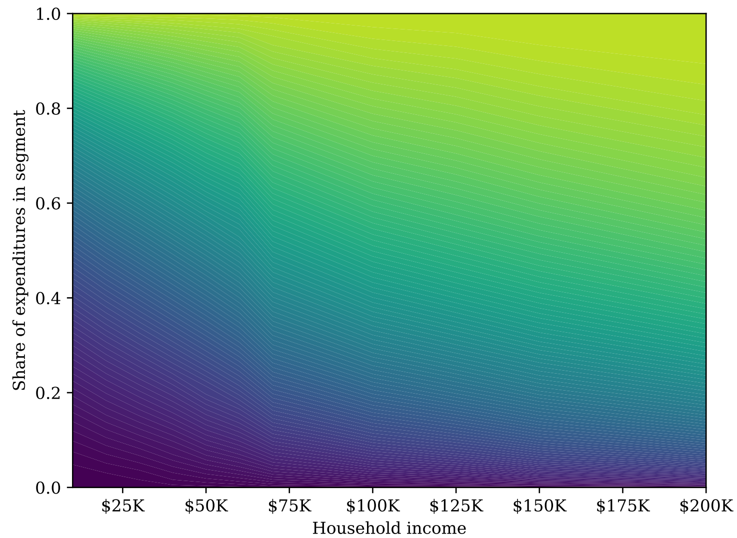


Figure F1: Expenditure shares χ_{ik}, using $K = 100$ groups.

We can use this definition of χ_{ik} to write the sales-weighted average markup paid by household i as

$$\mu_i^{avg} = \int_0^1 \chi_{ik} \frac{1}{1 - 1/\sigma_k} dF(k).$$

Calibration. I calibrate this version of the model in two steps. In step 1, I take expenditure shares directly from the data and choose elasticities of substitution by household income $\sigma(z_i)$ to match the average retail markups μ_i^{avg} paid by income group. In step 2, I use the elasticities of substitution and implied prices to back out how quality shifters differ by income group $q_k(z_i)$. I discuss each of these steps in more detail.

For step 1, I start by taking expenditure shares directly from the data. In principle, one could take expenditure shares for each UPC separately and solve the model this way. To make this calibration more transparent, I aggregate UPCs into K different segments. I do so by ordering UPCs from lowest to highest average buyer income (as in the extension of the model in the main text with heterogeneous products), and then splitting the UPCs into K segments with equal sales. I choose $K = 100$ and show the expenditure shares of each income group on the 100 product segments in Figure F1. As expected, low-income households have higher expenditure shares on lower-ranked UPC segments, while high-income households have higher expenditure shares on higher-ranked UPC segments.

I then fit the elasticities of substitution by household income $\sigma(z_i)$ to match the observed sales-weighted markups recorded in Figure 1. To reduce dimensionality, I impose that $\sigma(z_i)$ is a second-degree polynomial in log income.[58] Figure F2 plots the estimated elasticities of substitution by

[58]Specifying a third- or fourth-degree polynomial for this relationship does not significantly affect the results, though in those cases the elasticity of substitution for the highest income group is below one.

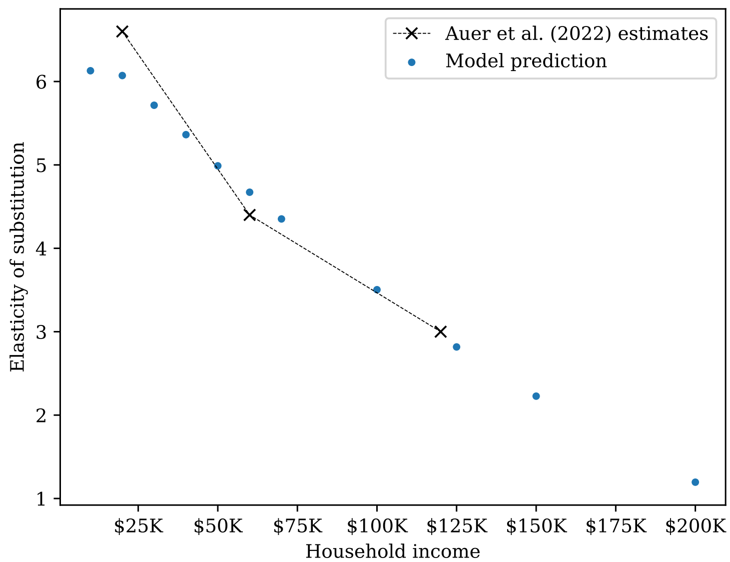


Figure F2: Calibrated elasticities of substitution $\sigma(z_i)$, compared to estimates from Auer et al. (2022).

income, with estimates from Auer et al. (2022) as a comparison. Elasticities of substitution decrease from about 6 for the lowest-income households to 1.2 for the highest-income households in the sample. These calibrated estimates line up closely with estimates from Auer et al. (2022).

For step 2, note that the ratio of expenditures by household i on two types k and k' is given by

$$\frac{p_k c_{ik}}{p_{k'} c_{ik'}} = \left(\frac{\mu_k}{\mu_{k'}}\right)^{1-\sigma_i} \left(\frac{A_{k'}}{A_k}\right)^{\sigma_i - 1} \left(\frac{q_{ik'}}{q_{ik}}\right)^{\sigma_i} = \left(\frac{\mu_k}{\mu_{k'}}\right)^{1-\sigma_i} \frac{\hat{q}_{ik'}}{\hat{q}_{ik}},$$

where the last equation uses an "adjusted quality-shifter" $\hat{q}_{ik} = A_k^{\sigma_i - 1} q_{ik}^{\sigma_i}$. Hence, I use the expenditure share of each income group on each product k and the markups μ_k implied by the elasticities of substitution to calculate $\hat{q}_k(z_i)$ for all z_i and all k.

To calculate how changes in the income distribution over time affect the aggregate retail markup, I hold $\sigma(z_i)$ and $\hat{q}_k(z_i)$ fixed and recompute expenditure shares and markups using the distribution of post-tax real income documented by Saez and Zucman (2019) from 1950 to 2018. Table F1 reports the results. Changes in the income distribution from 1950–2018 account for a 15.7pp rise in the aggregate retail markup. Changes in income dispersion contribute 2.6pp of the overall rise and about 20 percent of the rise since 1980.[59]

Compared to the calibration presented in the main text, this version of the model predicts a greater increase in the aggregate retail markup over time. Intuitively, this is because the search

[59] The relative role of the changes in the income level and changes in income dispersion is sensitive to the polynomial fit to $\sigma(z_i)$, because it depends on the concavity of σ in log income. For example, if we instead fit a third-order polynomial to $\sigma(z_i)$, the counterfactual exercise attributes 4.0pp of the predicted increase in the aggregate retail markup to changes in income dispersion.

Table F1: Predicted change in markup from 1950–2018 under calibration with heterogeneous elasticities of substitution and taste for quality.

Period	Predicted Δ agg. markup	Portion due to Δ Income level	Δ Income dispersion
1950–2018	15.7	13.1	2.6
1950–1980	4.0	3.8	0.2
1980–2018	11.7	9.3	2.4

model in the main text has an additional feedback loop. When the composition of households shifts toward higher-income households, aggregate search intensity falls, and since households' search decisions are strategic substitutes, each household individually exerts more search effort. In this version of the model, households' elasticities of substitution are static, and so individual price sensitivity does not increase as the economy shifts towards higher incomes.