

Markups Across the Income Distribution: Measurement and Implications

Kunal Sangani

Harvard University*

October 13, 2023

Download latest version

Abstract

I examine the relationship between customer income and firm markups using rich data on household transactions and wholesale costs. Over the observed purchases, high-income households pay 15pp higher retail markups than low-income households. Half of the markup gap is due to differences in markups paid at the same store. Conditional on income, markups paid by a household also increase when a household shops in high-income areas, at retail chains with locations in other high-income areas, or when purchasing products with a high-income customer base. A model in which household search intensity depends on opportunity cost of time can account for these facts. Consistent with the model's predictions, I document that retail markups across cities rise with both per-capita income and inequality. Through the lens of the model, changes in the income distribution since 1950 account for a 10–14pp rise in retail markups, with 25 percent of the increase due to growing income dispersion. The rise in markups consists of within-firm markup increases as well as a reallocation of sales to high-markup firms, which occurs without any change to the nature of firm production or competition.

*Email: ksangani@g.harvard.edu. First version: March 2022. I am grateful to Gabriel Chodorow-Reich, Andrei Shleifer, and Ludwig Straub for their excellent advice and support. This paper also benefited from comments by David Baqae, Adrien Bilal, Xavier Gabaix, Deivy Houeix, Erik Hurst, Kiffen Loomis, Alexander MacKay, Pierfrancesco Mei, Guido Menzio, Namrata Narain, Pascual Restrepo, Peleg Samuels, Lawrence Summers, Adi Sunderam, and participants at the NBER Summer Institute 2022 and SED 2023. 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

A growing body of evidence suggests that average markups in the U.S. economy are rising (De Loecker et al. 2020; Barkai 2020; Autor et al. 2020; Gutiérrez 2017). Many of the mechanisms put forward to explain this phenomenon attribute 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), or 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.¹ A large literature in industrial organization and trade finds that price sensitivity tends to decline with income (see e.g., Nevo 2001; Handbury 2021; Auer et al. 2022). If this is the case, a shift in the composition of demand toward high-income households should lead to a decline in aggregate price sensitivity and hence a rise in markups. Yet, the magnitude of this force and its contribution to the rise in markups are unclear.

This paper estimates the relationship between household income and firms' markups and explores how changes in the income distribution affect the evolution of markups over time. The first part of the paper is empirical: using rich data on household transactions and wholesale costs for fast-moving consumer goods, I quantify how retail markups vary with household income and with aggregate income across cities and over time. In the second part of the paper, I develop a macroeconomic model that can account for the evidence and use the model to simulate how retail markups evolve with changes to the income distribution.

I begin by constructing a dataset on retail markups—hereafter referred to as markups for brevity—by pairing transaction-level data on household purchases with data on wholesale costs faced by retailers. My use of price and cost data to measure retail markups follows Gopinath et al. (2011) and Anderson et al. (2018), who argue that, since rent, capital, and labor are fixed at short horizons, merchandise costs are a natural proxy for the marginal costs faced by retailers.² The merged dataset includes markups for over 26 million transactions made in a single year. Relative to using data from a single retailer,

¹Exceptions include Bornstein (2021), who studies how aging demographics affect firms' markups, and Döpper et al. (2021), who attribute rising markups in scanner data to a secular decline in price sensitivity. I discuss the relationship between my mechanism and Döpper et al. (2021) in the related literature below.

²Two other ways markups are measured in the literature are demand estimation and production function estimation. In Appendix E.2, I show that the retail markups in my dataset are strongly correlated with markups recovered from demand estimation in a selected product category (margarine). (Vertical conduct matters for whether demand estimation recovers retailer markups, manufacturer markups, or a combination; I briefly discuss these conduct assumptions in Appendix E.) Appendix E.1 shows that markups estimated using the production function approach by De Loecker et al. (2020) also increase with buyer income.

this merged dataset has two advantages: (1) since it includes all household expenditures in NielsenIQ-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.

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 44 percent for households with over \$200,000 in annual income. This 15pp markup gap is due to differences in markups paid for identical products (as previously documented by Aguiar and Hurst 2007 and Broda et al. 2009) as well as differences in basket composition: high-income households tend to buy products with higher average markups.³ To relax the assumption that local inputs and transport costs do not affect marginal cost, I also consider more conservative measures of the markup gap that control for the county and store of purchase. In both cases, a significant gap in markups paid by high- and low-income households remains: three-quarters of the markup gap persists within county, and just over half of the markup gap persists within store.⁴

While these estimates describe how markups vary with household income in the cross-section, how markups vary with aggregate income also depends on spillovers that are absorbed in the intercept of cross-sectional regressions (Nakamura and Steinsson 2014; Chodorow-Reich 2020; Wolf 2021). Across an array of specifications that exploit different sources of variation, I find positive and large spillovers of others' incomes on markups paid. For example, conditional on household income, a household pays higher markups in high-income counties, when purchasing products with a high-income customer base, and when shopping at retail chains with locations in other high-income areas. My preferred specifications identify these positive spillovers using variation over time in the income of a product or retail chain's customers, controlling for time-varying unobserved costs in the area where a household shops.

This positive dependence of markups paid on others' incomes implies that the "macro" elasticity of markups to income—that is, the relationship between markups and aggregate

³This paper is the first to quantify the effects of basket composition on markups paid across income groups. Previous work (e.g., Broda et al. 2009; Griffith et al. 2009; Handbury 2021) shows that households can save by substituting to lower-quality products or cheaper stores, but lack common units in which to compare products of different qualities or stores with different amenities. The use of cost data to construct markups provides a common unit for these cross-good comparisons.

⁴The within-store markup gap controls for store-specific factors that may affect marginal costs, but omits differences in markups across income groups that are due to shopping at stores with different average markups. Kaplan and Menzio (2015) show that differences in the expensiveness of stores where a household shops constitute a large share of the variation in price indices across households.

income—is larger than the “micro” elasticity observed in the cross-section. While the elasticity of markups to income in the cross-section of households is 3–4 percent, I estimate a macro elasticity of markups to income of 8–15 percent. Empirical designs that exploit variation in income in the cross-section and over time produce similar estimates.

I use these moments from the data to develop and discipline a macroeconomic model in which markups set by firms depend on the income distribution. The model features informational frictions that lead households to retrieve price quotes from firms and generates price dispersion for identical products in equilibrium, as in the canonical nonsequential search model of Burdett and Judd (1983).^{5,6} I add two key ingredients to the model that cause the income distribution to play a role in determining firms’ markups. First, I allow the distribution of price quotes retrieved by each household to depend on the household’s endogenous choice of search effort. Second, I allow households to differ in labor and search productivity, leading to heterogeneous opportunity costs of time and thus search intensities across households.⁷ As a result, households’ search decisions and the distribution of markups charged by firms are all endogenous, equilibrium outcomes that depend on the income distribution.

The model accounts for the patterns recorded in the data—high-income households pay higher markups, and positive spillovers across households generate a macro elasticity of markups to income greater than the micro elasticity—and produces additional predictions for search behavior that I test in the data. In particular, I show that search intensity in the data is decreasing in own income (as previously documented by Pytka 2018) but increasing in average county income, consistent with the prediction that household search decisions are strategic substitutes in equilibrium. In contrast, an alternate class of models in which differences in markups across income groups arise due to non-homothetic preferences or differences in utility primitives (e.g., Berry et al. 1995; Simonovska 2015; Auer et al. 2022) are unable to account for these relationships between search behavior and income or for the contribution of differences in prices paid for identical products in the same market to the markup gap across income groups.

I use the model to return to the question that opens this paper: how do changes in the

⁵ Appendix G shows that a sequential search model à la Burdett and Mortensen (1998) generates similar results. I contrast these search models against a model in which differences in markups paid across households arise due to differences in utility primitives (such as elasticities of substitution) in Appendix H.

⁶ Price dispersion in the model is “spatial” (across retailers), but since retailers are indifferent over a range of prices, can be interpreted to also include intertemporal price dispersion. In the data, ability to exploit spatial and intertemporal price variation is the dominant source of differences in prices paid for identical products across income groups (coupon usage plays a minor role in comparison).

⁷ Pytka (2018) and Nord (2022) also develop search models with heterogeneous households. I clarify the differences between my approach and theirs in the literature review below.

income distribution affect aggregate markups? I derive conditions in the model under which a rise in income levels or a rise in income dispersion lead to higher markups and show that both sets of conditions hold when I calibrate the model to match differences in markups paid across income groups. Simulating the model with the income distributions of different U.S. cities generates a positive relationship between markups and both income level and dispersion across cities, consistent with the data. Remarkably, the model's predicted markups across cities account for over 30 percent of the variation in aggregate markups across cities in the data, three times more than a representative agent, nested CES model that predicts markups using data on retailer market shares and concentration.

The model suggests that changes in the income distribution can play a meaningful role in the evolution of markups over time. I consider changes in the distribution of post-tax real income in the U.S. documented by Saez and Zucman (2019). Through the lens of the model, the increase in income levels and increase in income dispersion together account for a 10–14pp rise in the aggregate retail markup from 1950 to 2018.⁸ This increase is in line with data on retail gross margins reported by the Census of Annual Retail Trade Statistics. Increases in the aggregate markup are moderate from 1950–1980 but accelerate after 1980 due to rising income inequality.

In the model, half of the rise in the aggregate markup over this time period is due to a reallocation of sales to high-markup firms. The expansion of high-markup firms occurs because declining search intensity leads households to shop more often at higher quantiles of the markup distribution. The contribution of reallocations to the increase in the aggregate markup is consistent with evidence from De Loecker et al. (2020) and Autor et al. (2020). However, while some studies interpret the increased market share of high-markup firms as the proximate cause of increasing markups, in this paper both reallocations and rising markups are consequences of changes in the demand side of the economy. In all, the calibration suggests that changes to the income distribution can be a potent force in both reshaping market structure and increasing the aggregate markup over time.

Related literature. This paper's empirical findings on differences in markups paid across income groups are closely related to a literature that documents differences in prices paid for identical products, starting with seminal papers by Aguiar and Hurst (2007) and Broda et al. (2009). When costs are constant across retailers, differences in prices paid for identical products isolate within-product differences in markups paid. My use of wholesale cost data to construct retail markups facilitates comparisons across products (as well as across

⁸Rising markups do not necessarily imply rising profits. In the model, profits are static due to free entry.

periods when product costs change). I find that taking differences in basket composition into account doubles the elasticity of markups paid to household income compared to the within-product component alone, because the baskets of high-income households tend to be composed of higher-markup products.⁹

This paper also adds to a literature in macroeconomics and trade that provides evidence of a relationship between income (or wealth), price sensitivity, and markups, including Lach (2007), Gicheva et al. (2010), Alessandria and Kaboski (2011), Simonovska (2015), Anderson et al. (2018), Stroebel and Vavra (2019), DellaVigna and Gentzkow (2019), Handbury (2021), Faber and Fally (2022), and Auer et al. (2022).^{10,11} Most closely related to my study of the relationship between markups and aggregate income are Simonovska (2015), who documents how markups charged by a single e-commerce retailer vary with per-capita income across export destinations, and Anderson et al. (2018), who explore how markups within two retailers vary with local income across store locations. The dataset I construct allows me to capture patterns of substitution across retailers, as well as decompose the relationship between markups and aggregate income into partial and general equilibrium effects. Nevertheless, estimates from Simonovska (2015) and Anderson et al. (2018) are similar in magnitude to my estimates of the macro elasticity of markups to income (8–15 percent).¹²

A vast literature in industrial organization estimates markups in consumer markets and often finds a negative relationship between income and price sensitivity. For example, Nevo (2001), Villas-Boas (2007), Nakamura and Zerom (2010), and Grieco et al. (2021) find that high-income consumers are less price sensitive in each of the markets they study (breakfast cereals, yogurt, coffee, and automobiles). While the demand systems used in these studies can generate a relationship between income and markups, they attribute

⁹It is not obvious that this across-product component of the markup gap should be positive. In fact, classic models of second-degree price discrimination suggest that a monopolist may choose to set higher markups for products purchased by low-income households (see e.g., Mussa and Rosen 1978, Tirole 1988 Ch. 3, and references therein). In contrast, I find a positive relationship between markups and income across products sold by the same manufacturer and across products sold at the same retail outlet.

¹⁰To my knowledge, the relationship between income and price sensitivity was first conjectured by Harrod (1936), who called this relationship the Law of Diminishing Elasticity of Demand. A related marketing literature starting with Dickson and Sawyer (1990) and Hoch et al. (1995) also explores how income and other characteristics predict a household's likelihood to be "coupon-prone."

¹¹Differences in price sensitivity across income groups estimated in these studies differ widely. For example, Auer et al. (2022) estimate that elasticities of substitution fall with income from 6.6 to 3.0, while Faber and Fally (2022) estimate differences in elasticities of substitution across low- and high-income customers of 0.4. Direct evidence on markups paid by income groups helps discipline these differences. My estimates are similar in magnitude to Auer et al. (2022).

¹²Simonovska (2015) reports an elasticity of markups to per-capita income across export destinations of 12–24 percent. Anderson et al. (2018) find that the elasticity of gross margins to local income across store locations is 10 percent for a Canadian retailer and 17 percent for a U.S. retailer.

differences in price sensitivity to utility primitives and are unable to account for differences in prices paid for identical products in the same market, which I find constitute half of the markup gap across income groups. It is also difficult to infer a macroeconomic relationship between markups and income in these models because they are partial equilibrium and typically estimated within narrow product categories.

On the theoretical front, my model of consumer search builds on the insights of Aguiar and Hurst (2007) and Kaplan and Menzio (2015) that search and opportunity cost of time play an important role in prices paid.¹³ I build on the search technology developed by Burdett and Judd (1983).¹⁴ Pytka (2018) and Nord (2022) also develop models in which households have heterogeneous search intensities, and both embed these households in an Aiyagari (1994) model to explore differences between expenditure and real consumption inequality.¹⁵ 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 unique to this paper.

Finally, this paper relates to recent work that documents trends in markups over time and considers potential drivers (e.g., De Loecker et al. 2020; Barkai 2020; Gutiérrez 2017).¹⁶ Most closely related are Döpper et al. (2021) and Brand (2021), who explore how markups recovered from demand estimation in retail scanner data evolve over time. Both studies provide complementary evidence that demand-side forces may play an important role in the evolution of markups. Döpper et al. (2021) attribute rising markups in scanner data to declining consumer price sensitivity and incomplete pass-through of marginal cost reductions.¹⁷ Brand (2021) also finds that markups have increased and attributes the rise to consumers becoming less price sensitive, perhaps due to increased product

¹³Theoretical foundations for the importance of search in explaining price dispersion go back to Stigler (1961). 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).

¹⁴The Burdett and Judd (1983) technology has been used for a number of macroeconomic applications, such as understanding differences in export prices across countries (Alessandria and Kaboski 2011), the degree of vertical differentiation (Albrecht et al. 2021) and specialization (Menzio 2021), and relative price dispersion (Kaplan et al. 2019).

¹⁵Nord (2022), which was released after the first draft of this paper, also provides evidence that the skewness of price distributions in NielsenIQ data covaries with customers' incomes, consistent with a search model.

¹⁶Not all markets appear to follow a trend of increasing markups. For example, Berry and Jia (2010) find that markups fell in the airline industry from the 1990s to 2006 due to more elastic passenger demand, and Grieco et al. (2021) find that markups in the automobile industry fell from 1980 to 2018.

¹⁷Döpper et al. (2021) find that changes in buyer income are partly responsible for trends in markups over time across product groups, but conclude that a secular decline in price sensitivity plays an even larger role. One possibility is that this decline in price sensitivity is due to changes in the income distribution that are not picked up by noisy measures of buyer income. My model also provides a rationale for why search behavior within income level may shift over time as a strategic response to a changing income distribution.

differentiation. Neither of these papers ties the rise in markups to changes in the income distribution, as I do in this paper.

Layout. Section 2 describes the data and how I measure retail markups. Section 3 presents the main empirical findings. Section 4 develops a model in which differences in markups paid arise from differences in search intensity and basket composition. Section 5 shows that the model’s predictions on search behavior are borne out in the data. Section 6 calibrates the model, and Section 7 explores macroeconomic implications of the model, including spillovers across income groups, effects of inequality, and markups across cities. Section 8 considers how changes in the income distribution over time affect retail markups. Section 9 describes extensions developed in the Online Appendix, and Section 10 concludes.

2 Data Construction

In this section, I describe two data sources—NielsenIQ Homescan and PromoData Price-Trak—that I use to construct measures of retail markups paid by households. Appendix A provides a detailed description of the procedure for cleaning and merging the two datasets and describes other ancillary data sources used in the paper.

2.1 Data sources

Consumer panel data. NielsenIQ 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.¹⁸

NielsenIQ Homescan panelists use in-home scanners or a mobile application to record all purchases intended for personal, in-home use. The data include all NielsenIQ-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 NielsenIQ does not pay panelists, it offers households a variety of incentives to accurately report data.

¹⁸From 2006–2009, NielsenIQ separately identifies households with \$100K, \$125K, \$150K, and over \$200K in household income. These distinctions are not available before 2006 or after 2009. I report results for 2007 to avoid the impact of the recession in 2008–2009. Appendix Table B1 reports a subset of the results for other years.

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

For some analyses, I make use of NielsenIQ’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.¹⁹

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 2006–2012.²⁰ On a weekly basis, wholesalers send PromoData list prices and promotional discounts that they make available to their customers. Previous studies using these data include Nakamura and Zerom (2010), Stroebel and Vavra (2019), and Afrouzi et al. (2021).

The PromoData report both base prices and “deal prices.” Deal prices include promotional discounts and are only available to retailers during windows scheduled by the wholesaler. (They may also require retailers to provide proof of promotion in order to redeem the discounted price.) In the main text, I present results using deal prices as the measure of retailers’ wholesale costs. The results are not sensitive to instead using base prices as the measure of wholesale costs (see Table 1).

As previously documented by Stroebel and Vavra (2019), list wholesale costs from the PromoData 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 (see Appendix Table A1).²¹ Hence, for my baseline results, I calculate a national wholesale price for each UPC in each month. Similar results obtain when using only the subset of household transactions where a wholesale cost is reported for the market where the purchase is made (see Table 1).

In all, about 67,000 UPCs purchased by Homescan panelists in 2007 are matched to

¹⁹For example, the “jams, jellies and spreads” product group consists of nine product modules for jams, jelly, marmalade, preserves, honey, fruit spreads, peanut butter, fruit and honey butters, and garlic spreads.

²⁰A 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).

²¹Using cost data from a major grocer, DellaVigna and Gentzkow (2019) also find that wholesale costs across stores do not vary with local income (see DellaVigna and Gentzkow 2019 Appendix Figure 14).

wholesale costs from PromoData. These UPCs constitute 43 percent of transactions and 37 percent of expenditures in the 2007 panel.

2.2 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 ,

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

Here, wholesale costs in the month of purchase proxy for the replacement costs that retailers face for restocking items purchased by customers. As argued by Gopinath et al. (2011), these replacement costs serve as a reasonable measure of marginal costs since other components of retailers' costs, such as rent, capital, and labor, are fixed at short horizons. The approach of using data on retailers' replacement costs to measure retail markups follows Aguirregabiria (1999), Eichenbaum et al. (2011), Gopinath et al. (2011), and Anderson et al. (2018), among others.²²

While my baseline assumption is that these wholesale costs accurately measure retailers' marginal costs, the analysis below seeks to control for three ways in which list wholesale costs in the PromoData may mismeasure the marginal costs faced by retailers: (1) true wholesale costs may vary from the list prices recorded by PromoData due to retailer-wholesaler or retailer-manufacturer 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 shipping costs, and (3) true marginal costs may differ from replacement costs, for instance due to local inputs for shelving and inventory management. In Section 3, I discuss empirical strategies that aim to control for each of these potential sources of mismeasurement.

Retail markups are winsorized at the 1 percent level for all analyses. The cost-weighted average markup in the merged Homescan-PromoData dataset is 32 percent. This is comparable to an average markup of 28 percent in data from Dominick's Finer Foods and an average markup of 41 percent for grocery stores in the 2007 Census Annual Retail

²²Eichenbaum et al. (2011), Gopinath et al. (2011), and Anderson et al. (2018) each use data on replacement costs provided by a retailer, which include wholesale costs as well as shipping costs and net rebates. The PromoData include list wholesale costs, but not shipping costs or rebates, and hence I conduct robustness exercises below that control for unobserved store-level costs and exclude perishable items. Reassuringly, Gopinath et al. (2011) report that their results are qualitatively unchanged if they use wholesale costs as a proxy for replacement costs, and Anderson et al. (2018) report that gross margins calculated using replacement costs versus using costs of goods sold are closely correlated.

Trade Statistics.²³

Selection. One might be concerned that selection into the set of products with wholesale costs biases the results presented below. Appendix Table A3 shows that the share of transactions and expenditures that are matched to wholesale costs is similar across income groups. Appendix Table A3 also compares the relative unit prices of matched and unmatched products for each income group. Relative unit prices of products that are not matched to wholesale cost data exhibit a larger covariance with income than those in the matched sample, suggesting that differences in markups across income groups in the matched sample are likely to be conservative.

Comparison to other markup measures. Two other approaches used to measure markups in the literature are demand estimation and production function estimation. In Appendix E, I estimate a random coefficients model à la Berry et al. (1995) to recover marginal costs and markups in a single product category (margarine). The recovered markups exhibit a strong positive correlation ($\rho \approx 0.6$) with the retail markups constructed using wholesale cost data.²⁴ Since the NielsenIQ data do not include retailer names, it is more challenging to compare retail markups in my data with firm-level markups measured using the production function approach. Nevertheless, Appendix E shows that markups of public retail firms measured by De Loecker et al. (2020) also exhibit a positive relationship with buyer income, consistent with the findings in this paper.

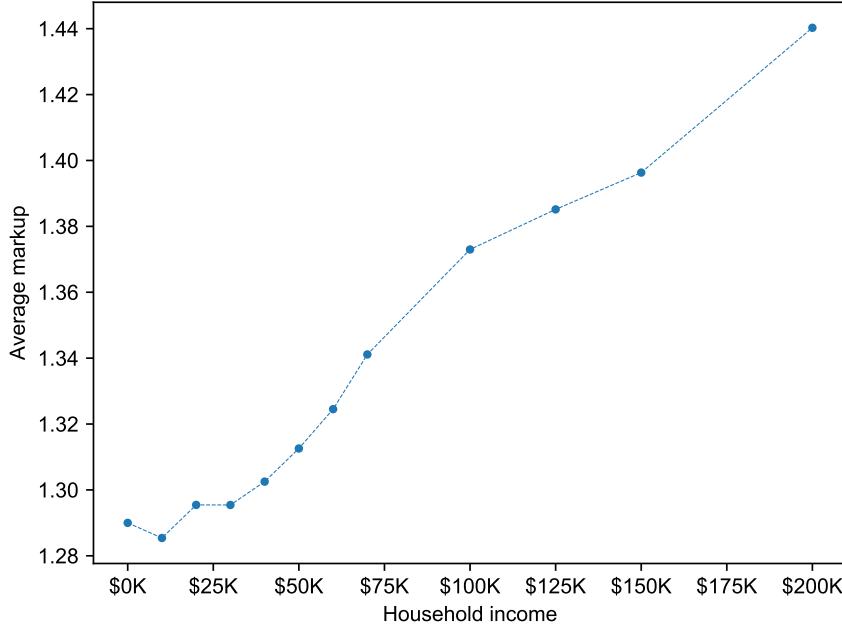
3 Empirical Evidence

In this section, I explore the relationship between markups and household income. Section 3.1 explores the relationship between markups and household income in the cross-section. High-income households pay higher markups on average. Differences in markups paid within store—which plausibly control for unobserved components of marginal costs such as labor or rent—account for over half of the gap. Section 3.2 documents that, conditional on income, markups paid by a household are also increasing in the income of other households. These spillovers contribute to the relationship between markups and aggregate income in an economy.

²³The 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 $1/(1 - 0.289) = 1.41$.

²⁴Whether markups recovered from demand estimation include retailers' markups depends on assumptions about vertical conduct between retailers and manufacturers, which I discuss in Appendix E.

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



3.1 High-income households pay higher markups

Figure 1 plots the aggregate (cost-weighted average) markup paid by households over the income distribution for the sample of purchases matched to wholesale costs. The aggregate markup increases from 29 percent for the lowest-income households in the sample to 44 percent for households with over \$200K in annual income.

To address the concerns discussed above about unobserved components of marginal cost, I document the difference in markups paid between low- and high-income households within county and within store. These controls absorb factors that may lead to systematic differences in marginal cost across counties or across stores, such as differences in shipping or local input costs, thus isolating the differences in markups.

The first specification adds demographic controls and county fixed effects:

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

The markup paid by household i for good g in transaction t is $\text{Markup}_{i,g,t}$.²⁵ 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

²⁵I use the retail markup in levels as the dependent variable for all regression results in the main text. Similar results obtain using the log retail markup as the dependent variable (see Appendix Table B1).

fixed effects; and $\epsilon_{i,g,t}$ is a mean-zero error. I weight the regression by costs and leave out the income level indicator for households with less than \$20K in income, so that the coefficients β_ℓ are differences relative to the group with below \$20K in income.

Figure 2a plots the coefficients β_ℓ for specification (1) with and without county fixed effects. After accounting for demographic controls, the fixed effect for the highest-income group is 18pp, slightly larger than the unconditional difference of 15pp. After adding county fixed effects, the fixed effect for the highest-income group falls to 14pp. Hence, about three-quarters 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.

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

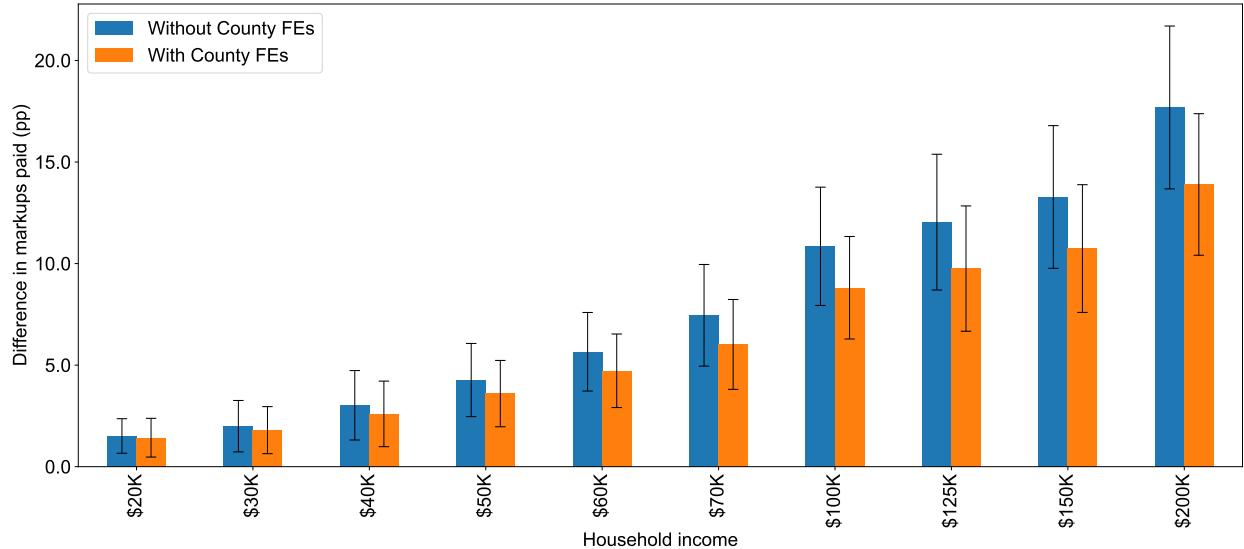
$$\text{Markup}_{i,g,t} = \sum_{\ell} \tilde{\beta}_{\ell} 1\{i \text{ has income level } \ell\} + \tilde{\gamma}' X_i + \alpha_{\text{Store}} + \epsilon_{i,g,t}. \quad (2)$$

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 gap in markups paid across the lowest- and highest-income groups is 13pp, which is moderately smaller than in the full sample.²⁶ Adding store fixed effects brings the difference in markups paid by the low and high-income households to 7pp. Accordingly, for the sample of transactions made at a NielsenIQ-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.

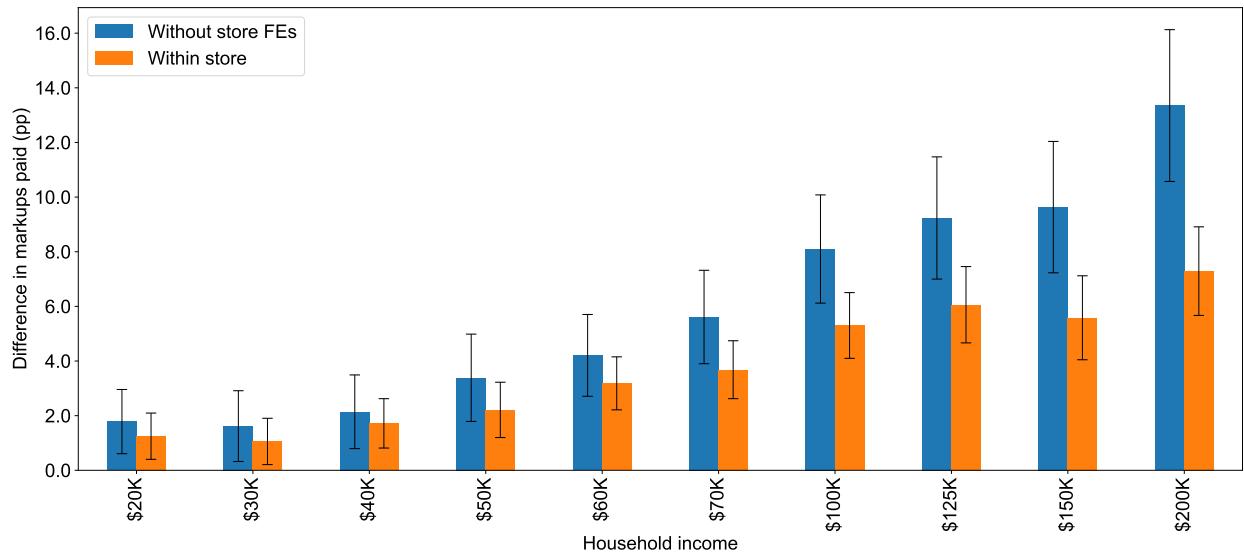
Why do households pay different markups even within the same store? Appendix Figure B2 shows that about half of the within-store markup gap is due to differences in basket composition—high-income households buy a sample of products that have higher average markups—and the other half is due to differences in prices paid for the same UPC at the same store. I return to these two sources of the markup gap below.

²⁶A review of the data suggests that NielsenIQ store IDs are only available for the subset of retailers that also participate in the NielsenIQ Retail 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 NielsenIQ-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 (identified store sample, $N = 14.0$ million).

Note: These figures plot the coefficients β_ℓ on household income dummies in a cost-weighted regression of markup paid on household income dummies and demographic controls (race, ethnicity, household size, and presence and age of female head of household). Income levels on the horizontal axis are the minimum of the income bracket provided by NielsenIQ. 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)).

3.1.1 Robustness

The markup gap across income groups is robust to concerns about unobserved volume discounts for large retailers and to using alternate measures of retailers' replacement costs.

Differential volume discounts by retailer size. While the store fixed effects in specification (2) absorb systematic differences in wholesale, shipping, and local input costs that cause marginal costs to differ by store, they do not absorb heterogeneity in marginal costs by product-store pair. This is a problem if some retailers face lower marginal costs on a subset of items. A salient concern is that large retailers may be able to negotiate volume discounts on a subset of products (e.g., on commodity items but not luxury items).²⁷ In this case, the data would overstate the marginal cost and underestimate 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 mismeasurement would lead me to overestimate the difference in markups paid across income groups.

I address this concern by testing whether the markup gap is driven by large retailers in the sample. I rank retailers by total sales and estimate specification (2) for the subsample of transactions excluding the largest retailer, the largest three retailers, the largest five retailers, and so on. If the markup gap is partially driven by mismeasurement of marginal costs at large retailers, then the gap should attenuate as large retailers are removed from the sample.

Appendix Figure B1 shows that the estimated markup gap is stable as large retailers are removed from the sample. This suggests that the estimated difference in markups paid is not a result of differential quantity discounts at large retailers.

Spoilage and congestion costs, and other measures of wholesale costs. Table 1 shows that the unconditional, within-county, and within-store markup gaps are also robust to a number of different specifications. For example, the markup gap is similar if regressions are weighted by sales rather than costs; if the PromoData base price, which excludes promotional discounts, is used as the measure of wholesale costs; and if the markup gap is measured using only the subset of transactions for which PromoData reports a wholesale cost of the purchased item in the market of purchase. Including day-of-week fixed effects, which control for changes in congestion faced by retailers over the course of the week, or supply-side controls, which include sales shares of the purchased UPC and

²⁷This concern is most salient with respect to retailer size, since the Robinson–Patman Act limits wholesalers and manufacturers from selling an identical product to different retailers at different prices, but does permit differences in prices that reflect differences in cost of delivery or volume.

Table 1: Robustness of markup gap.

Markup gap (pp) relative to <\$20K	Demographics		Within County		Within Store	
	\$100K	\$200K	\$100K	\$200K	\$100K	\$200K
Baseline	10.8	17.7	8.8	13.9	5.3	7.3
Weighting by sales	10.4	17.8	7.8	12.7	4.5	6.1
Using PromoData base price	9.7	16.0	8.0	12.7	5.6	7.9
Using PromoData market-level price	8.5	16.3	6.6	11.7	5.8	8.9
With day-of-week fixed effects	10.6	17.5	8.6	13.8	5.1	7.1
With supply-side controls	10.6	17.4	8.5	13.6	5.2	7.2
Excluding perishable items	10.5	17.1	8.4	13.4	4.8	6.7

Note: Results from fixed effect specifications (1) and (2) under alternate measures. The baseline measure is the cost-weighted gap in markups constructed using deal prices from PromoData.

brand and sales concentration in the product module, also do not alter the results. Finally, limiting the sample to non-perishable items, and thus excluding food items that may have higher shipping and spoilage costs, does not materially affect the results.

Appendix Table B1 also reports the share of the markup gap that persists within counties and within stores for a number of alternate specifications and for years 2006, 2008, and 2009. Results from all these alternatives are quite similar.

3.1.2 Accounting for the markup gap

Two forces contribute to the gap in markups paid across income groups. First, high-income households pay higher prices for identical products. This fact has been documented in seminal papers by Aguiar and Hurst (2007) and Broda et al. (2009). Differences in markups paid for identical products account for about half of the markup gap conditional on demographics, within county, or within store (see Appendix Figure B4).²⁸ In principle, differences in prices paid for identical products could arise from differences in exploiting variation in posted prices or from differences in coupon usage. In the data, coupons play a negligible role (Appendix Figure B3 shows that savings due to coupon usage vary by less than 1pp across income groups), suggesting this component of the markup gap is predominantly due to differences in how different income groups exploit spatial and intertemporal variation in posted prices.

Second, the remainder of the markup gap is due to differences in the composition

²⁸That is, the gap in markups paid for identical products between the highest and lowest income groups data is 9pp conditional on demographics, 7pp within county, and 4pp within store. Compare these to the overall markup gaps of 18pp conditional on demographics, 14pp within county, and 7pp within store.

of shopping baskets across income groups. High-income households also pay higher markups on average because their baskets contain a larger share of high-markup products. This paper is the first to estimate this source of markup differences across income groups. Broda et al. (2009), Griffith et al. (2009), and Handbury (2021) show that households can save by substituting to lower-quality products or cheaper stores, but lack common units in which to compare products of different qualities or stores with different amenities. The use of cost data to construct markups provides a common unit for cross-good comparisons.²⁹

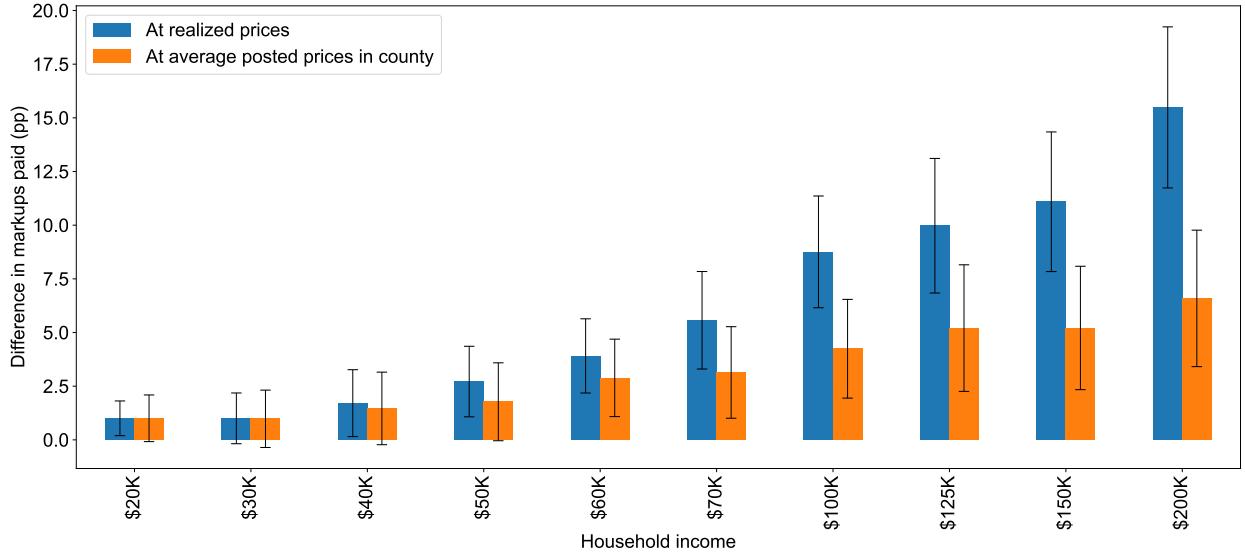
Including this second source of markup differences across households results in an elasticity of markups to household income that is two times larger than previous estimates that only exploit differences in prices paid for identical products. Appendix Table B2 shows that the elasticity of markups paid to household income is 3.2–3.8%, compared to estimates of the elasticity of prices paid for identical products to income of 1.1–1.3% in Broda et al. (2009).³⁰

Some differences in basket composition across income groups could be due to differences in search behavior (for instance, if some products exhibit systematically higher variation in prices). To decompose the markup gap into a component due to search and a component due to basket composition, I construct a measure of markups that each household would pay if it instead paid the average posted price for a UPC in its county, using data on posted prices from the NielsenIQ Retail Scanner dataset. These counterfactual markup measures remove any of the effects of consumer search from markups paid, thus isolating effects of basket composition. Figure 3 plots the overall markup gap across households against the markup gap that would result if households paid the average posted price for each UPC in their county. These estimates suggest basket composition constitutes about 40 percent of the markup gap. I return to these moments when calibrating the model in Section 6.

²⁹This across-product component of the markup gap could also be positive or negative in theory. For example, classic models of quality discrimination, such as Mussa and Rosen (1978) and Tirole (1988) Ch. 3, suggest that firms may want to set low markups for high quality goods purchased by high-income customers in order to deter high-income customers from substituting to lower quality goods. The model developed in this paper instead predicts that average markups on goods purchased by high-income households are higher due to differences in buyer composition, consistent with the data.

³⁰Replicating the elasticity of prices paid for identical products to household income in my sample produces slightly higher estimates of 1.7–2.0% (see Appendix Table B2). The estimates in Broda et al. (2009) may be lower because Broda et al. (2009) are not able to disaggregate households with over \$100K income into separate income groups, because of differences in the sample year (Broda et al. 2009 use data from 2005, while I use data from 2007), or because of minor differences in the set of demographic controls used. Results are similar if we instrument for household income using household heads' education and occupation, to address concerns about noise in reported income or idiosyncratic income shocks (see Appendix Table B3).

Figure 3: Counterfactual markup gap if households instead paid average posted prices within county for each UPC.



Note: The figure plots coefficients from a cost-weighted regression of markup paid on income dummies. The blue bars use markups constructed from households' transaction prices, while the orange bars use counterfactual markups constructed from the average NielsenIQ Retail Scanner posted price for the product in the county of purchase. Standard errors are two-way clustered by product brand and household county.

3.2 Markups depend positively on other buyers' incomes

This section explores how markups paid by a household depend on the income of other buyers. These spillovers are important for characterizing how the *macro elasticity* of markups to income—how the economy's aggregate markup changes if all household incomes change—relates to the micro elasticity of markups to income observed in the cross-section. To a first order, the macro elasticity $\varepsilon_{\mu, \text{Income}}^{\text{agg}}$ is the sum of this micro elasticity and the elasticity of markups to others' incomes:

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

If retailers 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, the cross-elasticity $\varepsilon_{\mu, \text{Others' Incomes}}^{\text{individual}} = 0$, and the macro elasticity of markups to income coincides with 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.³¹

³¹As I discuss at the close of this section, the cross-elasticity may be positive or negative in theory. For

Figure 4: Average retail markup paid by income group, split by (a) quintile of county income and (b) average income of other UPC buyers.

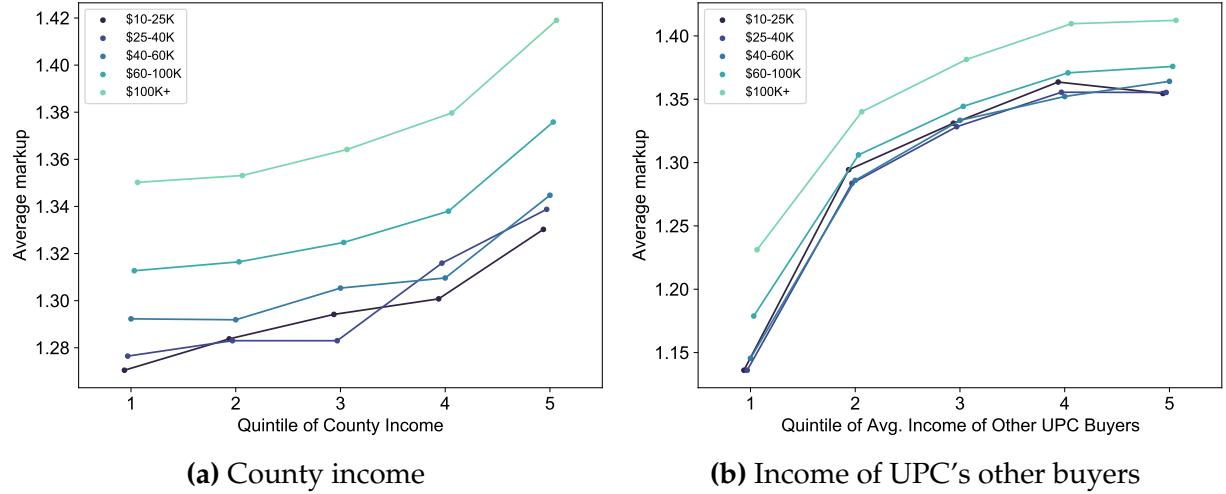


Figure 4 shows the average retail markup paid by households in five income groups split by (a) quintile of county income and (b) quintile of the average income of a UPC’s other buyers. In both figures, consistent with the findings from the previous section, the average markup paid by the highest income group lies above the average markup paid by the lowest-income group in all cases. Conditional on income, markups paid also increase with county income and with the income of other buyers of the same product. These patterns suggests that markups paid by a household increase when the household is “pooled” with higher-income shoppers.

Of course, this descriptive plot does not account for potentially unobserved components of marginal cost or correlations between unobserved household characteristics and location/product choice. To sharpen identification, I exploit time series variation in the incomes of buyers that a household shops alongside. For this purpose, I compile data on retail markups for all years from 2006 to 2012 (over which I have data on wholesale costs from PromoData). I consider three different definitions of the pool of shoppers that a household is grouped with, each of which exploits a different source of variation and allows different controls.

The first specification considers how markups paid by household i for UPC g at store s in year y on transaction t depend on average income in the core-based statistical area

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.

(CBSA) where the household lives.³²

$$\log \text{Markup}_{i,g,s,y,t} = \beta_1 \log \text{CBSA Income}_{\text{CBSA}(i),y} + \gamma_{i,\text{IncomeLevel}(i,y)} + \alpha_s + \delta_y + \varepsilon_{i,g,s,y,t}. \quad (3)$$

Here, the fixed effects $\gamma_{i,\text{IncomeLevel}(i,y)}$ absorb household characteristics that may be correlated with a household's choice of location as well as changes in income level within a household. The store fixed effects α_s absorb potentially unobserved components of marginal cost that vary at the store level but are fixed over time, and the year fixed effects δ_y absorb secular time trends in unobserved components of marginal cost. Hence, this specification is intended to isolate variation in markups paid due to variation in average CBSA income over time. To measure average CBSA income, I use annual estimates from the Bureau of Economic Analysis (BEA).

Table 2 column 1 reports the results from specification (3): the elasticity of CBSA markups to per-capita income is 7.1 percent. Note that these spillovers are inclusive of endogenous responses by households and firms, such as entry of new stores, changes in how households allocate spending across stores and goods, and so on.

The second specification considers how markups paid depend on the income of other buyers at the same retail chain:

$$\begin{aligned} \log \text{Markup}_{i,g,s,y,t} = & \beta_2 \log \text{Income at Retailer Locations}_{\text{Retailer}(s),y} \\ & + \gamma_{i,y} + \alpha_s + \phi_{\text{County}(i),y} + \varepsilon_{i,g,s,y,t}. \end{aligned} \quad (4)$$

To calculate average income at a retailer's locations, for each store s , I take the sales-weighted average of per-capita county income (from the BEA) across all stores in the same retail chain and same designated market area (DMA) as store s .

Note the addition in (4) of county-year fixed effects $\phi_{\text{County}(i),y}$, which absorb variation in county-level costs over time, and of household-year fixed effects $\gamma_{i,y}$, which absorb changes in household income or other characteristics over time. These additional controls overcome concerns from (3) that the estimated elasticity of markups to others' incomes may be biased by unobserved changes in marginal costs within an area that are correlated with changes in local income or by changes in unobserved household characteristics. The fact that two stores in the same area in the same year can effectively be exposed to customer bases with different incomes—due to changes in customer income at other stores in the same retail chain—allows for these additional controls.

³²CBSAs refer collectively to metropolitan and micropolitan statistical areas in the U.S., and consist of multiple counties. These are more disaggregated than combined statistical areas (CSAs).

Table 2: Spillovers of other buyers' incomes on markups paid, exploiting variation over time from 2006–2012.

<i>Log Retail Markup</i>	(1)	(2)	(3)
Log CBSA Income	0.071** (0.013)		
Log Income at Retailer's Locations		0.065** (0.027)	
Log Income of Other UPC Buyers			0.142** (0.038)
Household × Income Level in Year FEs	Yes		
Year FEs	Yes		
Household-Year FE		Yes	Yes
Store FE	Yes	Yes	
County-Year FE		Yes	
Store-Year FE			Yes
N (millions)	91.9	50.8	97.0
R ²	0.19	0.21	0.21

Note: The sample includes all household transactions matched with wholesale cost data from 2006 to 2012. Regression weighted by sales (in 2007 USD), and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

The intuition for the variation exploited in specification (4) is best summarized by the following experiment. Suppose household i shops at two retail chains A and B. Retailer A has locations in another city C_A and retailer B has locations in another city C_B . If incomes in city C_A rise relative to C_B , how do the markups paid by household i at retailer A change compared to at retailer B? If retailers are able to perfectly price discriminate across customers, we should not expect changes in income in other areas to affect the markups paid by a household. However, frictions that prevent retailers from price discriminating perfectly across customers—such as the patterns of uniform pricing documented by DellaVigna and Gentzkow (2019) and Hitsch et al. (2021)—could lead to spillovers across households.

Column 2 of Table 2 reports that the elasticity of markups paid to average income across a retailer's locations is 6.5 percent. That is, if the average income across retailer A's locations increase 10 percent relative to average income across retailer B's locations, the markups paid by a household at retailer A's store increases 0.65 percent relative to markups paid by the same household at retailer B's store.

Finally, the third specification exploits variation in the income of buyers of the same

UPC over time,

$$\log \text{Markup}_{i,g,s,y,t} = \beta_3 \log \text{Income of other UPC buyers}_{g,y} + \gamma_{i,y} + \psi_{s,y} + \varepsilon_{i,g,s,y,t}, \quad (5)$$

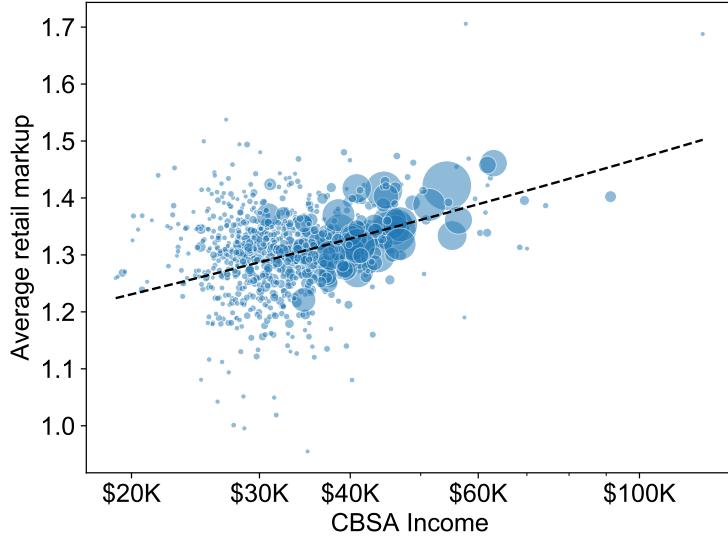
where the store-year fixed effects $\psi_{s,y}$ control for potentially time-varying unobserved components of cost at the store level. I calculate the income of the other buyers of UPC g in year y as the sales-weighted average income of all households purchasing g in year y excluding household i . Table 2 column 3 reports an elasticity of markups paid to the income of a UPC's other buyers of 14.2 percent.

Thus, across all three specifications, Table 2 finds positive and large spillovers of others' incomes on markups paid, ranging from 6 to 14 percent. These spillovers are economically meaningful, since the elasticity of markups paid to own income is 2.1–3.8 percent. Combining the micro elasticity of markups to own income with the elasticity to others' incomes, my preferred estimates for the macro elasticity of markups to income fall between 8 and 15 percent.

Robustness. In Appendix Table B5, I show that the sign and magnitude of these spillovers is robust to various changes in the estimating equations. Analogous specifications to (3)–(5) can also be estimated in a single cross-section of data, though using a cross-section of data prevents us from including store fixed effects to absorb store-level variation in costs. As shown in Appendix Table B4, similar patterns emerge in the cross-section: conditional on own income, households pay higher markups when shopping in high income CB-SAs, when shopping at retail chains with other high income customers, or when buying products with a high-income customer base.

Heterogeneity across income groups. Appendix Table B11 explores whether the direction and magnitude of spillovers are heterogeneous across income groups. In the Varian (1980) model of sales, for example, increasing the share of uninformed buyers (taken as a proxy for price-insensitive, high-income households) increases prices paid by uninformed buyers but decreases prices paid by informed buyers (taken as a proxy for price-sensitive, low-income households), since informed buyers benefit from the entry of more stores. I find that spillovers are positive for households in all income groups, with marginally larger spillovers for high-income households. I return to this evidence when discussing pro-competitive effects in the model in Section 4.8.

Figure 5: CBSA per-capita income and average retail markup.



Note: Each bubble is a core-based statistical area (CBSA). The size of each bubble is proportional to total CBSA expenditures in the NielsenIQ Homescan data. CBSA per-capita income in 2007 is from the BEA. The dashed line is from a regression of log markup on log per-capita income, weighted by CBSA sales.

Comparison to existing estimates. While this paper is the first to estimate and decompose the aggregate relationship between markups and income into own-income effects and spillovers, we can compare the macro elasticity of markups to income to other estimates in the literature. For example, using data from a single online apparel retailer, Simonovska (2015) finds that the elasticity of markups charged on identical products to per-capita income across export destinations is 12–24 percent. Anderson et al. (2018) calculate the elasticity of gross margins to local income across store locations within two retailers and arrive at estimates of 10 and 17 percent. Using the opening of large plants as a natural experiment, Bhardwaj et al. (2022) find an elasticity of local retail prices to changes in hourly wages of about 10 percent. My estimates of a macro elasticity of markups to income of 8–15 percent are broadly in line with all three previous findings.

This macro elasticity is also in line with the relationship between average markups and per-capita income across CBSAs in the data. Figure 5 plots aggregate (cost-weighted average) retail markups in each CBSA against per-capita income. The elasticity of CBSA markups to income is 11 percent, squarely in the range of macro elasticities estimated using the specifications (3)–(5) that account for potentially unobserved local costs.

Why doesn't the elasticity of markups to aggregate income, as measured in the cross-section of CBSAs, seem to be biased by unobserved local costs? In Appendix D.1, I use data from the 2007 Census Retail Trade Survey of Detailed Operating Expenses along with

data on retail wages and rents to estimate the potential bias in the cross-CBSA elasticity due to unobserved local costs. This analysis suggests that even if a substantial portion of labor and rent expenses are recategorized as variable costs (instead of overhead costs), the bias in the measured elasticity is mild. This is because retailers' total labor and rent costs are small relative to costs of goods sold.

Macro elasticities across space vs. over time. The final set of analyses in this section explores whether the link between markups and aggregate income across space and over time differ. A concern is that if different spillovers operate across space (e.g., due to uniform pricing by retailers operating across multiple cities) compared to over time (e.g., due to entry of new products/stores or technological improvements), then a model that accounts for the relationship between markups and aggregate income in one dimension may perform poorly for counterfactuals in the other.

To answer this question, I estimate the relationship between aggregate markups and income across CBSAs in both the cross section and time series. Table 3 columns 1–2 estimate the elasticity of markups to CBSA income using data from 2006 to 2012, over which wholesale costs from PromoData are available. Compared to column 1, which captures variation in incomes across CBSAs, column 2 adds CBSA fixed effects, thereby exploiting only variation in incomes over time within-CBSA. The two estimates are nearly identical, suggesting that the link between markups and income over time within each CBSA is similar to the relationship between markups and income across CBSAs.

Since this exercise is limited to a short sample period, I conduct a similar analysis using data on unit prices paid by households in the NielsenIQ Homescan data from 2004–2019. Here, average unit prices (e.g., the average price paid per ounce of margarine) within a product category serve as a proxy for markups.³³

These analyses again suggest similar cross-sectional and time series relationships. Columns 3 and 5 isolate cross-sectional relationships between unit prices and income, by estimating how average unit prices for a product module in a given year vary with CBSA income, controlling for average differences in unit prices across CBSAs in all years. On the other hand, columns 4 and 6 isolate the time series relationships between unit prices and income, by comparing how average unit prices for a product module in a given CBSA vary with income over time, controlling for shifts in unit prices in all product categories across years. Whether looking across all years from 2004 to 2019 or only looking at changes over from the beginning to end of the sample (and thus removing the influence of business

³³As in Handbury (2021), I exclude any product modules where product sizes are expressed in counts, so that unit prices (per unit of weight or volume) are a meaningful basis for comparison.

Table 3: Cross section vs. time series elasticities of markups to aggregate income.

Outcome Sample	<i>Log Markup</i>		<i>Log Avg.</i>		<i>Unit Price</i>	
	2006–2012		2004–2019		2004 and 2019	
	(1)	(2)	(3)	(4)	(5)	(6)
Log CBSA Income	0.104** (0.017)	0.099** (0.026)	0.127** (0.026)	0.128** (0.030)	0.097** (0.045)	0.096* (0.052)
Year FEs	Yes	Yes		Yes		Yes
CBSA FEs		Yes	Yes		Yes	
Year-Product Module FEs			Yes		Yes	
CBSA-Product Module FEs				Yes		Yes
<i>N</i> (millions)	133	133	18.3	18.3	2.2	2.2
<i>R</i> ²	0.03	0.04	0.99	0.99	0.99	0.99

Note: The sample for columns 1–2 includes all household transactions matched with wholesale cost data from 2006 to 2012. Columns 3–6 use quantity-weighted average unit prices for each product module in each CBSA, using NielsenIQ Homescan data from 2004–2019. Regression weighted by sales (in 2007 USD), and standard errors two-way clustered by CBSA and year. * indicates significance at 10%, ** at 5%.

cycles), the elasticities across space and over time are nearly identical.

4 A Search Model of Income and Markups

In this section, I develop a quantitative model of consumer search and firm pricing that can account for the empirical patterns documented in the previous section. Besides generating additional testable predictions, the model allows us to explore more ambitious counterfactuals—such as how changes in income dispersion affect aggregate markups—that are difficult to isolate in the data.

Households in the model exhibit heterogeneity along two key dimensions. First, households with different income levels have heterogeneous tastes, which results in differences in basket composition across income groups. Second, households endogenously choose different search intensities. The latter draws on a large literature that documents the importance of search effort in prices paid (e.g., Aguiar and Hurst 2007, Alessandria and Kaboski 2011, Kaplan and Menzio 2015).

In partial equilibrium, differences in search intensity across households determine differences in markups paid for identical products. In general equilibrium, differences in search intensity, combined with differences in buyer composition across products, result in different average markups across products. Together, these forces generate both patterns documented in the previous section: (1) markups increase with household income, and

(2) the macro elasticity of markups to income is greater than the micro elasticity in the cross-section.

While the model rationalizes the empirical evidence presented in the previous section, one could consider different search technologies than the one I use here, or different micro-foundations altogether. Online Appendices G and H explore two alternatives, and I compare these approaches at the end of this section.

4.1 Households

A unit measure of households purchases goods indexed by $k = 1, \dots, K$. Household i 's utility is given by the CES preferences,

$$u(\{c_{ik}\}) = \left(\sum_{k=1}^K (\beta_{ik} c_{ik})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

where c_{ik} is household i 's consumption of good k and β_{ik} is a taste shifter for good k that is allowed to vary flexibly across households. In the data, these taste shifters will allow me to flexibly match differences in the composition of shopping baskets across income groups.

When purchasing each good k , households face an information friction common in canonical models of consumer search: households know the distribution of prices posted by firms, $F_k(p)$, 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 buying good k prior to purchase is a random variable with probability mass function $\{q_{ik,n}\}_{n=1}^\infty$. That is, with probability $q_{ik,1}$ the household observes only a single price quote, with probability $q_{ik,2}$ the household observes two price quotes, and so on. Each price quote retrieved is an independent draw from the distribution of prices F_k .

Upon receiving n price quotes, households compare the minimum price quote received to an exogenous 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.³⁴

The distribution of the number of price quotes received by household i is determined by the household's search intensity when buying good k , which I denote s_{ik} . Formally,

³⁴I assume each unit of consumption is infinitesimal, so that integer constraints can be ignored.

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 of $\{q_n\}_{n=1}^\infty$. I assume that $Q_1(s)$ is strictly decreasing in s , with $Q_1(0) = 1$ and $\lim_{s \rightarrow \infty} Q_1(s) = 0$. Additionally, I assume $Q_n(s)$ is weakly decreasing and C^∞ in s for all n and all s . These assumptions guarantee that if a household spends no time searching, it receives exactly one price quote, and that increases in search intensity lead to a first-order stochastic shift in the number of price quotes received.

Households choose time spent working, as well as consumption and search intensity for each good, to maximize utility subject to a time constraint and budget constraint:

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

where $t_i(c, s)$ is the time it takes household i to shop for c units with search intensity s , l_i is time spent working with labor productivity z_i , and p_{ik} is the average price paid by i per unit of consumption of good k . With infinitesimal units of consumption, there is no uncertainty in the average price paid p_{ik} or total consumption c_{ik} . 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 household i to shop for c units with search intensity s is

$$t_i(c, s) = \frac{c}{a_i} s, \quad (6)$$

where a_i is household i 's search productivity. Search productivity is allowed to vary to reflect the fact that households may differ in their 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.

Equation (6) assumes time spent shopping for a good increases linearly with the amount of consumption c and with search intensity s and decreases with search productivity a_i . The assumption that time spent shopping is linear in search intensity s is without loss of generality, since we could accommodate a different increasing relationship between shopping time and search intensity through of change-of-variables and an according change in the mapping function \mathcal{S} . On the other hand, the relationship between shopping time and the amount of consumption is more important. I make this assumption to reflect the observation in Pytko (2018) that shopping for a larger consumption basket takes more time. Appendix Tables B8–B9 provide additional evidence that shopping time

increases with basket size in the data.³⁵

Under utility maximization, households increase search intensity s_{ik} as long as the marginal benefit from increasing search intensity is greater than the marginal cost of doing so. Thus,

$$-\frac{\partial p_{ik}(s_{ik}, F_k)}{\partial s_{ik}} \leq \phi_i, \quad (7)$$

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 $s_{ik} = 0$. For the remainder of the text, I focus on the case where each household has an internal solution for s_{ik} in each market, so that (7) 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 the aggregate consumption of good k by C_k and the consumption-weighted distribution of labor productivity by

$$d\Lambda_k(z) = \frac{c_k(z)}{C_k} dH(z). \quad (8)$$

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

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

$$\bar{q}_{k,n} = \int_0^{\infty} q_{k,n}(z) d\Lambda_k(z), \quad \text{for all } n. \quad (9)$$

³⁵ Appendix Table B8 shows that household shopping time, proxied using measures from Kaplan and Menzio (2015), increases with total expenditures, controlling for household income, demographic characteristics, and the average markup paid by the household. Using household size as an instrument for total expenditures yields similar results. Appendix Table B9 shows that the positive relationship between basket size and shopping time also holds within-households over time. Note that while (6) assumes shopping time increases linearly in consumption, one could consider an elasticity of shopping time to consumption less than one. In the calibration, returns to scale in shopping for a larger total consumption basket are absorbed by differences in search productivity a_i calibrated across households.

This is a probability mass function because $q_{k,n}(z)$ sums to one over all n for each z and $d\Lambda_k(z)$ 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

Each good k is supplied by a measure M_k of ex ante identical firms, which 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 $\pi(p)$,

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

where the demand curve $D_k(p)$ that a firm faces depends on its price, the distribution of prices charged by other firms F_k , 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

$$F(p) = \begin{cases} 0 & \text{if } p < \underline{p} \\ 1 - \Psi\left[\left(\frac{R-1}{p-1}\right)\bar{q}_1\right] & \text{if } \underline{p} \leq p \leq R \\ 1 & \text{if } p > R \end{cases} \quad (10)$$

where the reservation price R is the highest price in the support of F , the lowest price \underline{p} in the support of F is

$$\underline{p} = 1 + \frac{\bar{q}_1}{\sum_{n=1}^{\infty} n\bar{q}_n}(R - 1), \quad (11)$$

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 supplying each good, M_k .³⁶ In equilibrium, the zero profit condition yields

$$\pi(p) - f_e = 0 \quad \text{for all } p \in [\underline{p}, R]. \quad (12)$$

4.3 Equilibrium

An equilibrium is a tuple $(F_k, \{c_k(z), s_k(z)\}_{z=0}^\infty, M_k)_{k=1,\dots,K}$ such that (1) consumption and search intensity chosen by households for each product, $c_k(z)$ and $s_k(z)$, maximize utility; (2) the distribution of prices for good k , F_k , is a dispersed price equilibrium given aggregate search behavior $\bar{q}_{k,n}$; (3) the mass of firms supplying each good k , M_k , is such that firms make zero profits net of the entry cost; and (4) all resource constraints are satisfied.

When search intensities are chosen endogenously, the Burdett and Judd (1983) framework produces two dispersed price equilibria, where one dispersed price equilibrium is stable and the other is unstable (see Burdett and Judd 1983 Theorem 2 and the graphical illustration in Appendix Figure C1). I conduct all comparative statics locally around the stable equilibrium in each market.

4.4 Characteristics of the search mapping S

It will be useful for the following results to define two additional conditions on the mapping S from search intensity to the distribution of price quotes received.

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

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

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

Note that the left-hand side of the expression in Assumption 1 is the second derivative of the expected price paid at search intensity s with respect to search intensity. Thus, Assumption 1 imposes that price is a decreasing and convex function in search intensity. In other words, returns to search diminish as search intensity increases.

³⁶Given a distribution of buyers' incomes, whether I assume free entry or an exogenous mass of firms does not affect the level of markups in the model. Appendix C shows that consumer search behavior pins down markups, and the free entry condition is cleared by changes in the mass of firms M .

Assumption 2. For any non-degenerate distribution F , the mapping \mathcal{S} satisfies

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

where Q_n and $\mathbb{E}[p|n]$ are as defined in Assumption 1.

Assumption 2 guarantees that when $\phi(z)$ 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 . This assumption is not necessary for most of the results that follow, but will be important when characterizing how a change in income dispersion affects markups.

While these two conditions may appear technical, I show in Appendix C.6 that both assumptions are satisfied by the two most common parameterizations of the Burdett and Judd (1983) model used in the literature: (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 is drawn from a Poisson distribution (e.g., Albrecht et al. 2021 and Menzio 2021).

4.5 Search and markups paid in the cross-section

Lemma 1 shows that how search intensity and markups paid vary over the income distribution depends on how the opportunity cost of search effort, $\phi(z)$, varies with income.

Lemma 1 (Search intensity and markups paid). *Suppose Assumption 1 holds. If the opportunity cost of search effort $\phi(z)$ is increasing (decreasing) in z , then*

1. *Search intensity $s_k(z)$ for any good k is decreasing (increasing) in z ,*
2. *Average prices $p_k(z)$ and markups paid $\mu_k(z)$ for any good k are increasing (decreasing) in z .*

Whether high-income households pay higher or lower markups for identical goods depends on whether search productivity $a(z)$ increases more or less than one-for-one with labor productivity z . This means that the model can in principle 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” (e.g., Caplovitz 1963; Prahalad and Hammond 2002); (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 for identical products.

While the empirical evidence presented in the previous section implies that the latter case is the relevant one in our setting, Lemma 1 explains why the relationship between

markups and customer income may vary in other settings. For example, Grunewald et al. (2020) find that low-income customers pay higher markups on average in the auto loan market, where consumer education and internet availability likely play a heightened role in consumers' abilities to gather and compare price quotes.³⁷

4.6 Strategic interactions in search

A household's choice of search intensity depends both on the household's cost of search effort and on the savings that can be generated by increasing search intensity. Since the prices set by firms—and thus the returns to search—are determined by aggregate search intensity, household search decisions are affected by the search behavior of other households. Lemma 2 shows that, when costs of search are sufficiently low, household search decisions are strategic substitutes.

Lemma 2 (Strategic substitutes in search). *There exists some ϕ^{cutoff} such that, if $\phi(z) < \phi^{cutoff}$ for all z , then (1) a stable dispersed-price equilibrium exists, and (2) in that equilibrium, household search decisions are strategic substitutes ($\partial s_{ik}/\partial s_{jk} \leq 0$ for any two households $i \neq j$ in market k).*

Intuitively, whether household search decisions are strategic complements or substitutes depends on whether an increase in search intensity by one household increases or decreases returns to search for other households. When search costs are low—and hence aggregate search intensity is high in the initial equilibrium—increases in aggregate search intensity decrease returns to search. Thus, when one household increases search intensity, other households decrease search intensity in response.³⁸

In Section 5, I show that search behavior in the data conforms with the predictions of Lemma 2. The link between aggregate search intensity and returns to search also provides a rationale for findings from Nevo and Wong (2019) that increases in household search effort during the Great Recession coincided with declines in the returns to shopping effort. As we will see, the strategic substitutabilities in search behavior moderate the effect of changes in the average cost of search effort on aggregate markups. Of course, this moderating feedback loop is absent in models where household price sensitivity is determined by exogenous utility primitives and thus unaffected by changes in the composition of households in the economy.

³⁷In Pytka (2018) and Nord (2022), high-income households instead search less due to a convex disutility of shopping time. As a result, these models cannot admit a "poverty premium" in some markets.

³⁸When $\phi(z)$ is too large for all z , no households can be incentivized to search, and the dispersed-price equilibrium ceases to exist. There are intermediate values of ϕ for which the dispersed-price equilibrium can exist and household search decisions are strategic complements. However, the strategic substitutes behavior in Lemma 2 is the relevant region both in the calibrated model and in data on search behavior.

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

Define the *aggregate markup* $\bar{\mu}$ as the ratio of total sales to total variable costs, and define $\bar{\mu}_k$ as the ratio of total sales to total costs for good k . Lemma 3 shows that the fraction of households receiving only one price quote in equilibrium is a sufficient statistic for the aggregate markup.

Lemma 3. *In equilibrium, the aggregate markup for good k is $\bar{\mu}_k = 1 + (R - 1)\bar{q}_{1,k}$. The aggregate markup in the economy is*

$$\bar{\mu} = 1 + (R - 1) \frac{\sum_k C_k \bar{q}_{k,1}}{\sum_k C_k}.$$

Intuitively, since firms selling good k must make identical profits at all prices in the support of F_k , and the only customers of a firm charging the highest price R are those that receive no other price quotes, the fraction of customers receiving only one price quote $\bar{q}_{k,1}$ pins down the profits of all firms and hence the aggregate markup for good k . The aggregate markup in the economy is then a cost-weighted average over all goods.

I will now consider how changes to the income distribution affect the aggregate markup. To provide an analytic characterization, I specialize to the case where $K = 1$, so that there is a single distribution of buyers' incomes $\Lambda(z)$.³⁹ Proposition 1 provides sufficient conditions for a first-order stochastic shift in $\Lambda(z)$ to increase the aggregate markup, and Proposition 2 does the same for a mean-preserving spread (increase in inequality).

Proposition 1 (First-Order Shift). *Suppose $K = 1$ and Assumption 1 holds, and let $\Lambda(z)$ be the initial distribution of buyers' incomes in equilibrium. Consider a perturbation in the household distribution such that the new $\tilde{\Lambda}(z)$ first-order stochastically dominates $\Lambda(z)$. This perturbation leads to an increase in the aggregate markup if $\phi(z)$ is increasing in z .*

Proposition 2 (Mean-Preserving Spread). *Suppose $K = 1$ and Assumptions 1 and 2 hold, and let $\Lambda(z)$ be the initial distribution of buyers' incomes in equilibrium. Consider a perturbation in the household distribution such that the new $\tilde{\Lambda}(z)$ is a mean-preserving spread of $\Lambda(z)$. This perturbation leads to an increase in the aggregate markup if $\phi(z)$ is increasing and convex in z .*

Propositions 1 and 2 follow from Lemma 3. 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 as each household modifies its search intensity) do not overwhelm the direct response.

³⁹Equivalently, I could allow $K > 1$ but assume that preferences are Leontief ($\sigma = 0$) and taste shifters $\beta_k(z) = \beta_k$ are identical across households. These assumptions ensure that the distribution of buyers' incomes $\Lambda_k(z) = \Lambda(z)$ is identical across goods.

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.

Corollary 1 (Balanced Growth). *Suppose $K = 1$. The aggregate markup is constant if either*

1. *Search productivity is linear in income, $a(z) = \alpha z$.*
2. *For each household i in the economy, $z'_i = \gamma z_i$, and $a'_i = \gamma a_i$.*

In the first case in Corollary 1, the opportunity cost of search effort is degenerate at $\phi(z) = \phi = 1/\alpha$, and so all households pay identical markups, and changes in the income distribution have no effect on the aggregate markup in the economy. In the second case in Corollary 1, opportunity costs of search effort differ across households, but by increasing labor productivity and search productivity at the same rate for all households, 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.

4.8 Discussion

Pro-competitive effects. Jaravel (2019) and Handbury (2021) find that increasing the share of high-income households in an economy leads to lower relative prices on goods consumed by high-income households. Their results are consistent with larger markets leading to a reduction in relative prices, either because larger markets lead to a reduction in marginal costs (e.g., due to economies of scale), because pro-competitive effects in larger markets compress markups, or both. Using price and cost data from a single retailer over a short time series, Jaravel (2019) concludes that both falling costs and markups play a role.

The model can be extended to account for these pro-competitive effects by adding an endogenous response of search productivity to the mass of firms supplying a good. That

is, for each household i buying good k , let search productivity follow

$$a_{ik} = \bar{a}_i M_k^\zeta.$$

Intuitively, as more retailers enter to supply good k , it is less costly for households to search across stores and get many quotes, and this endogenous improvement in search productivity puts downward pressure on the distribution of markups charged by firms. The elasticity of search productivity to the mass of firms, ζ , parameterizes the strength of this effect. I show in Appendix Figure B15 that a value of $\zeta = 0.3$ can replicate the elasticities of prices to market size documented in Jaravel (2021).

These pro-competitive effects do not appear to play a strong role in determining empirical patterns in the dataset of retail markups that I construct, however. Appendix Table B11 shows that, both in the cross-section of cities as well as in the time series, markups paid by high-income households if anything rise faster with aggregate income than markups paid by low-income households. A value of $\zeta > 0$, on the other hand, would imply that the elasticity of markups paid to aggregate income would be lower for high-income households, due to offsetting pro-competitive effects in goods purchased by high-income households.⁴⁰ Hence, my baseline results assume $\zeta = 0$.

Alternative models. Online Appendix G develops a version of the model that builds upon the model of sequential search by Burdett and Mortensen (1998), instead of the non-sequential 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 H instead posits non-homotheticities in elasticities of substitutes and taste for quality, following Handbury (2021) and Faber and Fally (2022). While that version of the model can be calibrated to match differences in retail markups paid across households, it does not account for differences in prices paid for identical products across income groups nor the patterns in search behavior that I document below in Section 5. Since price sensitivity in that model is determined by exogenous utility primitives rather than an endogenous search decisions, that model also lacks a moderating feedback loop between aggregate income and price sensitivity, and thus yields counterfactual results when used to predict markups across cities and over time.

⁴⁰The discrepancy in results could be explained by my focus on retail markups, rather than price indices, across cities (compared to Handbury 2021) and my use of difference-in-differences across cities to discipline these spillovers (compared to the aggregate time series evidence in Jaravel 2021). On the latter, Sangani (2023) cautions that inflation inequality over short samples may be influenced by commodity price movements, which have differential effects on the markups of products consumed by low- and high-income households.

5 Evidence on Search Behavior

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

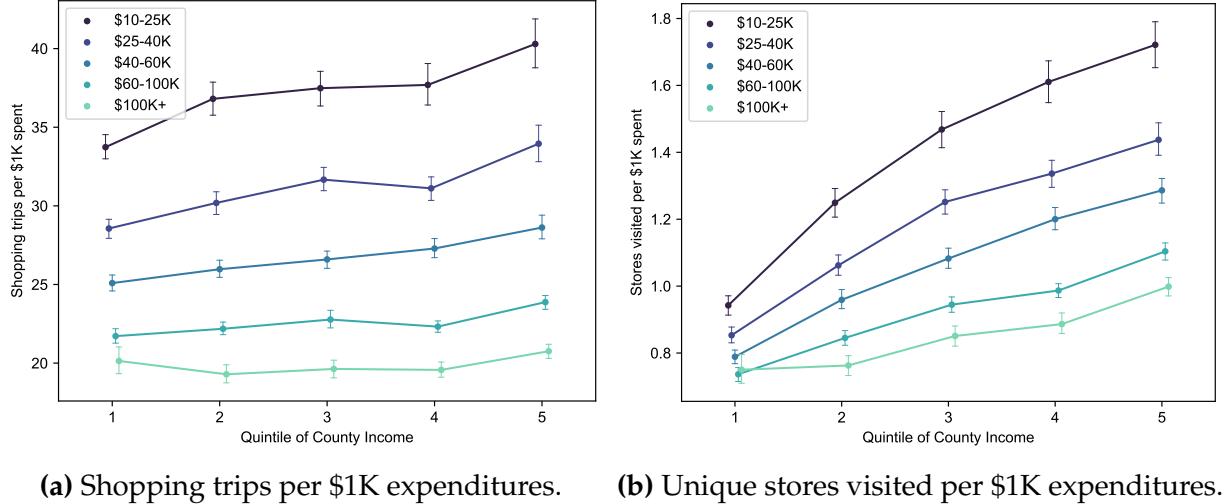
I employ two measures of shopping behavior introduced by Kaplan and Menzio (2015): the number of shopping trips a household makes and the number of unique stores it visits.⁴¹ I normalize both measures by dollar spent to account for the fact that total search time reflects both search intensity and the size of the consumption basket; my objective is to isolate the former. Of course, using expenditures to control for the size of the consumption basket risks confounding our results with differences in prices paid by income. Appendix Table B7 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 6 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 horizontal 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. Second, conditional on income, households exert greater search intensity in high-income counties.

Table 4 formally tests this relationship. Columns 1 and 4 report that shopping intensity is 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 searches more when surrounded by higher-income households.

⁴¹By making more shopping trips, a household is better able to detect and exploit intertemporal variation in prices. However, differences in the number of shopping trips across households may also reflect other factors, such as liquidity constraints. This motivates looking at several proxies for search effort.

Figure 6: Shopping intensity by income group and county income.



(a) Shopping trips per \$1K expenditures.

(b) Unique stores visited per \$1K expenditures.

Table 4: Effect of own 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 (IV)	-0.40** (0.01)	-0.40** (0.01)	-0.40** (0.01)	-0.35** (0.01)	-0.34** (0.01)	-0.34** (0.01)
Log Avg. County Income		0.19** (0.02)	0.09** (0.02)		0.65** (0.07)	0.26** (0.07)
Log Grocery Estabs.			0.03** (0.00)			0.11** (0.01)
State FEs		Yes	Yes		Yes	Yes
County FEs	Yes			Yes		
N	63 350	62 865	62 830	63 350	62 865	62 830
R ²	0.14	0.09	0.09	0.23	0.11	0.14

Note: Household income is instrumented using the education and occupation of the male and female heads of household. Average county income is from the BEA, and Grocery Estabs. are a count of NAICS 445 establishments from Census Business Patterns (includes grocery stores, supermarkets, liquor stores, and specialty food stores). Standard errors clustered by county. * indicates significance at 10%, ** at 5%.

6 Calibration

I calibrate the model developed in the previous section to match observed differences in basket composition across income groups and differences in markups paid across income groups. The calibrated model is able to replicate several untargeted moments in the data and from the literature.

6.1 Calibration procedure

There are two key differences across households with different income groups in the model: households have different preferences (e.g., different tastes), resulting in differences in basket composition across income groups, and households have different opportunity costs of search effort. I discipline household preferences using data on spending shares of different income groups across goods, and I discipline opportunity costs of search using data on average markups paid by income group.

Calibrating preferences across goods. In principle, one could calibrate the model by taking each UPC in the data as a distinct good.⁴² To reduce computational costs, I aggregate UPCs into $K = 10$ groups. I do so by ordering all UPCs in the data from lowest average buyer income to highest average buyer income and splitting the UPCs into K groups with equal sales. Appendix Figure B7 shows the fraction of expenditures by households at each income level on UPCs in each of the K groups for different values of K . As expected, low-income households disproportionately purchase UPCs in the lowest groups, while high-income households disproportionately buy from the highest groups. For the results reported in the main text, I choose $K = 10$, since increasing K beyond 10 does not materially change the results. (Appendix Table B13 reports results for different values of K ranging from 1 to 100.)

I choose preferences across income groups to exactly match these spending shares taken from the data. I assume preferences over goods are Cobb-Douglas ($\sigma = 1$) and choose each taste shifter, $\beta_k(z)$, to match the share of expenditures by households with income z on good k .

Note that the choice of Cobb-Douglas preferences across goods is without loss of generality for matching spending shares across income groups, since, for any other value

⁴²While the baseline calibration assumes each UPC is a distinct good, an alternative is to think of product categories as distinct goods, with consumers searching for items in that product category both across stores and across brands/products within a store. Previous work indicates that consumers face substantial cognitive costs when comparing products even in the same store (e.g., Seiler and Pinna 2017; Hossain 2020).

Table 5: Calibration parameter values and sources.

Parameter		Value	Source
Number of products	K	10^\dagger	Increasing $K > 10$ does not change results
Elasticity of substitution	σ	1^\dagger	Cobb-Douglas
Quality shifters	$\beta_k(z)$	-	Match spending shares
Unit wage	w	1	Numeraire
Reservation price	R	3.3^\dagger	98th percentile of markups in the data
Search mapping	\mathcal{S}	Poisson	Albrecht et al. (2021), Menzio (2021)
Opp. costs of search	$\phi(z)$	-	Match avg. markup paid by income group
Search productivity	$a(z)$	-	Solved from $\phi(z) = z/a(z)$
Household distribution	$H(z)$	-	Saez and Zucman (2019)

Note: \dagger indicates that the Online Appendix reports results under alternate values.

of σ , we could adjust the taste shifters $\beta_k(z)$ accordingly to continue matching the spending shares in the data. However, the assumption of Cobb-Douglas preferences across goods will matter when performing counterfactuals, since these preferences will determine how spending shares across goods will evolve as relative prices that households face for each good change. I show in Appendix Figure B16 that choosing other values for σ has very little effect on the results.

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

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

I take firms' unit costs as the numeraire and set households' reservation price to the 98th percentile of markups in the data, $R = 3.3$. Appendix Table B15 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.

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 in Figure 1. The calibration proceeds in two steps.

1. **Inner loop: Price distributions F_k and search intensities $s_k(z)$.** For each good k , the price distribution, F_k , and the search behavior of households shopping for that good, $s_k(z)$, solve a fixed point. Given an initial guess for $s_k(z)$, the price distribution F_k is pinned down by (10). The price distribution F_k in turn determines the returns to search

as a decreasing function in s_k . Setting returns to search equal to $\phi(z)$ yields a new estimate for households' search intensities, $s_k(z)$. I iterate the process until the price distribution F_k and search intensities $s_k(z)$ each converge.

2. Outer loop: Opportunity costs of search effort $\phi(z)$ and search productivities $a(z)$.

The outer loop calculates $\phi(z)$. Given an initial guess $\phi^{(0)}(z)$, I calculate the price distributions F_k and search intensities $s_k(z)$ for all k . Given F_k and the expenditure shares of households with income z across the K goods, there is a one-to-one mapping between opportunity cost of time and the average markup. Thus, I re-compute $\phi^{(1)}(z)$ to match the average markup paid by each income group. I iterate this process until the calibrated values $\phi^{(t)}(z)$ converge. Finally, I calculate search productivities using $a(z) = z/\phi(z)$.

The final parameter needed is the distribution of households across income groups. I use estimates from Saez and Zucman (2019), who report pre- and post-tax income by percentile of the income distribution from 1950 to 2018 (plotted in Appendix Figure B6). 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.

6.2 Calibrated statistics

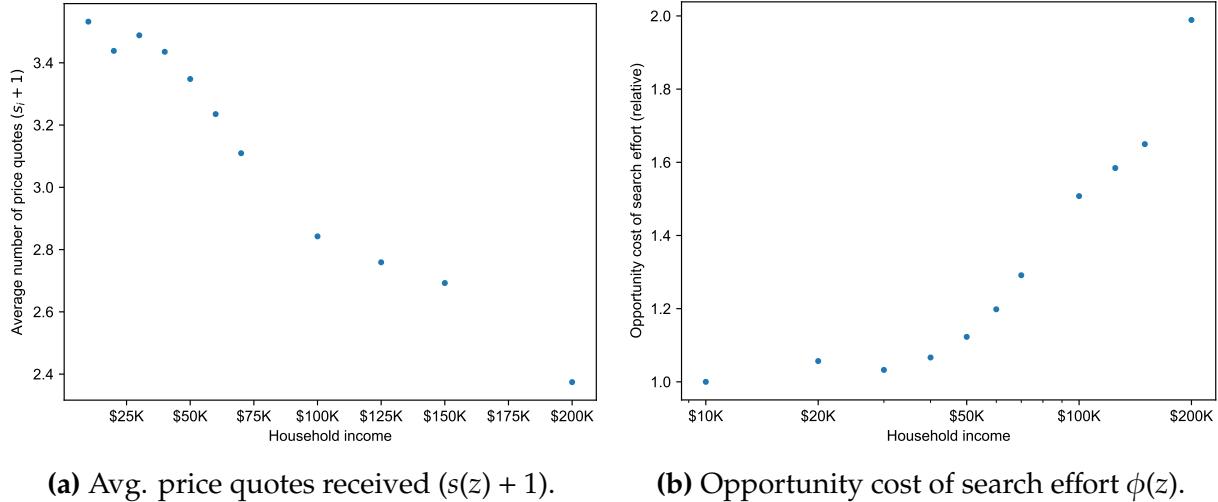
Figure 7 plots calibrated statistics on search intensity and opportunity cost of search effort by income group. The left panel shows that the expected number of price quotes, $(s(z)+1)$, decreases with income: households with over \$200K in income receive on average 30 percent fewer price quotes per purchase than low-income households. The calibrated opportunity costs of search effort $\phi(z)$, shown in the right panel, are increasing and convex in log income, satisfying the conditions in Propositions 1 and 2.

6.2.1 Untargeted moments

The calibrated model is able to account for several patterns in the data, including the relative contributions of search and basket composition to the markup gap across income groups, spillovers of other buyers' incomes on markups paid and search intensity, and the elasticity of price savings to search time.

Contributions of search and basket composition. We can isolate the relative contributions of search and basket composition to the markup gap across households in the model

Figure 7: Price quotes received and opportunity cost of search effort.



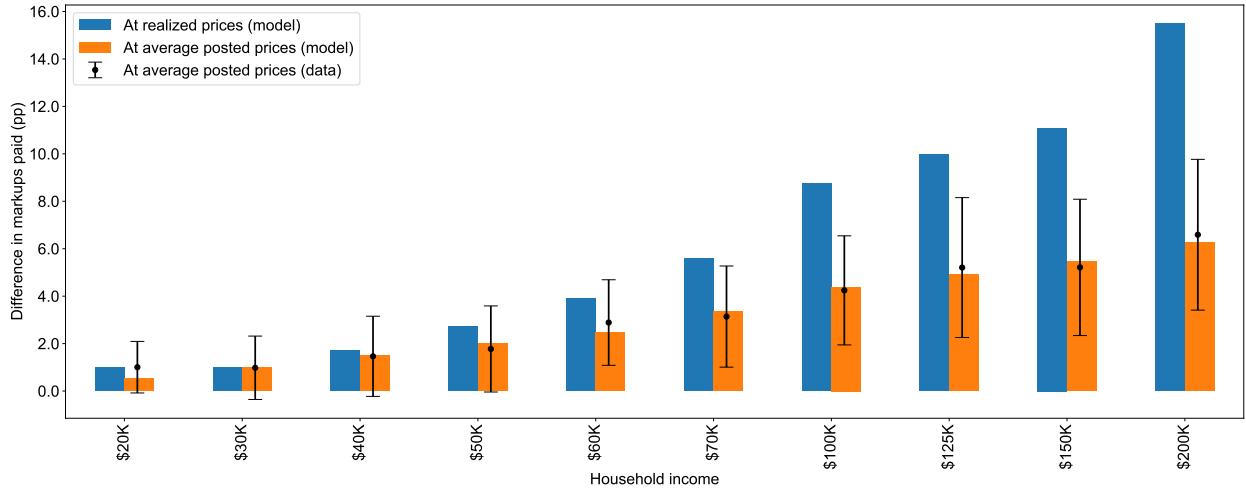
by calculating the markup gap across households that would result if households instead paid average posted prices for each good. This decomposition is analogous to the one constructed in the data in Figure 3.

Figure 8 plots the overall markup gap across income groups and the counterfactual markup gap that would result if all households instead paid average posted prices for each good k . The overall markup gap across income groups in the model matches the data by construction, but the counterfactual markup gap from paying average posted prices is an untargeted moment. The model is able to match the contributions of search and basket composition from the data quite closely. As an additional check, Appendix Figure B8 shows that the differences in prices paid for identical products across income groups in the model and the data are closely aligned.

Spillovers and strategic interactions. Table 6 compares the elasticity of markups paid and search intensity to own and others' incomes in the model to the data. To calculate how markups and search intensity vary with others' incomes, I simulate the model using the income distributions of 881 CBSAs—this exercise is described in detail in Section 7—and run regressions of markups paid and quotes received on own income and average income analogous to those in the data.

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 quite closely. The elasticity of markups paid to others' incomes, which is untargeted, is in line with the empirical evidence on spillovers from Section 3.2. As a result, the model produces a

Figure 8: Contributions of search and basket composition to markup gap: Model vs. data.



Note: The blue bars plot differences in markups paid by each income group relative to households with below \$20K income in the model. The orange bars plot differences in markups across income groups if all households instead paid the average posted price for each good. Empirical counterparts are from Figure 3.

macro elasticity of markups to income of 10.3 percent, squarely in the range estimated in the data (8–15 percent).

The elasticity of the average number of price quotes received to income in the model is -0.11 , within the range of -0.08 to -0.40 estimated in the data (see Appendix Table B7). The extent of differences in search intensity between high- and low-income households is also consistent with previous work by McKenzie and Schargrodsy (2005), who find that shopping frequency for households at the ninetieth percentile of the income distribution in Argentina is 30 percent lower than that of households at the tenth decile, after 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 strategic interactions in search documented in Section 5.

Other moments. In a seminal paper, Aguiar and Hurst (2007) estimate returns to shopping effort, finding that doubling shopping frequency lowers prices paid by 7–10 percent. I calculate this statistic in the model, and find that doubling time spent shopping decreases prices paid by 7.5–9.2 percent. (The range corresponds to differences across low- and high-income households.) In Appendix Table B12, I also compare various percentiles of the distribution of markups paid in the data to those predicted by the model for households at different income levels. While there are some differences of note—for example, house-

Table 6: Elasticity of markups paid and search intensity to own and others' incomes in the data and in the model.

	Log markup			Search intensity		
	OLS	IV	Model	OLS	IV	Model
Log Own Income	0.032	0.054	0.032	-0.26	-0.40	-0.11
Log Others' Income	0.102	0.089	0.071	0.03	0.09	0.03

Note: The first and second columns are from Appendix Table B4 columns 1–2, and the fourth and fifth columns are from Appendix Table B7. The “Model” columns report results from simulations of income distributions in 881 CBSAs, using avg. price quotes received per purchase as a proxy for search intensity.

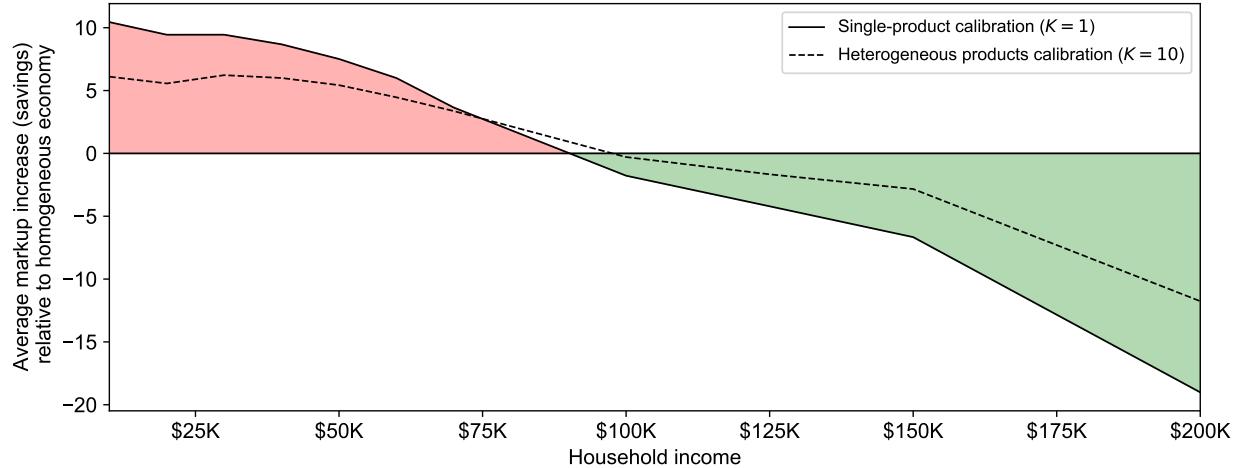
holds in the data make some purchases at markups below one, while the model does not admit prices below marginal cost—the model appears to fit the empirical markup distributions paid by each income group reasonably well.

6.2.2 Comparison to previous estimates of price sensitivity

Prior work estimating differences in price sensitivity across households (e.g., Faber and Fally 2022, Handbury 2021, Gupta 2020, and Auer et al. 2022) often estimate elasticities of substitution. The empirical evidence in this body of literature has been mixed, with some studies reporting small differences in elasticities of substitution across households (e.g., Faber and Fally 2022 find that differences in elasticities of substitution across the lowest- and highest-income groups is only 0.4) and others reporting large differences (e.g., Auer et al. 2022 estimate that elasticities of substitution decrease from 6.6 for low-income households to 3.0 for high-income households).

Differences in markups paid over the income distributions provide an additional moment that can discipline how large differences in price sensitivity across income groups must be. To compare estimates from the search model to estimates of elasticities of substitution from the literature, I calculate the aggregate markup that would result in an economy with homogeneous households of each income level and compute the elasticity of substitution $\eta(z) = \mu^{\text{homog}}(z)/(\mu^{\text{homog}}(z) - 1)$ that generates this markup. These elasticities of substitution fall from 5.5 for low-income households to just under 3.0 for high-income households, broadly aligning with the range of elasticities estimated by Auer et al. (2022) (see Appendix Figure B9).

Figure 9: Losses (savings) in markups paid relative to homogeneous income economies.



7 Cross-Sectional Implications

The calibrated model can be used to understand the nature of spillovers across households, to explore how changes in income dispersion affect markups paid by households across the income distribution, and to predict how markups vary across cities.

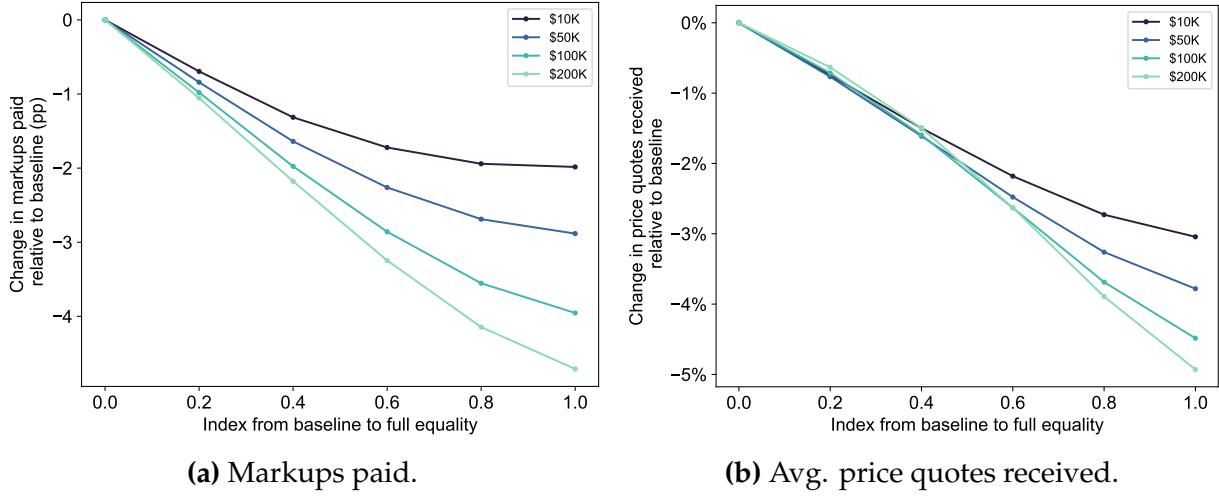
Spillovers. In the model, households' search decisions affect the distribution of prices offered by firms and hence affect the markups paid by other households. To quantify these spillovers, for each income level I consider an economy in which there is a representative household with labor productivity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

Figure 9 shows that the difference in average markups paid are large in magnitude: low-income households would pay 5pp lower markups in an economy populated with only low-income households. On the other hand, the highest-income households save over 10pp on markups paid due to the greater search intensity of low-income households. These spillovers are responsible for the wedge between the micro and macro elasticities of markups to income.

The composition of households in the economy affects not only markups paid, but also search behavior. Low-income households retrieve 6 percent more price quotes on average in the baseline 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.

Figure 9 also shows that these spillovers are larger in a single-product calibration

Figure 10: Effects of changing income inequality on markups and search behavior.



of the model. Intuitively, in the model with heterogeneous goods, differences in basket composition across income groups result in partial market segmentation, and thus dampen spillovers across income groups.

Effects of income dispersion. To explore the effects of inequality on households at different income levels, I consider how markups paid and search time would change if reduced income inequality. To do so, I simulate the model under counterfactual household distributions that scale all household incomes toward the average per-capita income in the economy.⁴³ Figure 10 plots markups paid and search intensity by four income groups as we change income dispersion from its baseline level to full equality. Reducing income dispersion leads to a decline in markups paid and in search intensity for all income groups. The intuition for this result goes back to the convexity of opportunity cost of search time in income: as income dispersion in an economy increases, even holding average income constant, the decline in search intensity by high-income households is greater than the increase in search intensity by low-income households. Hence, aggregate search intensity falls and markups rise for all households.

Interestingly, the effects of reducing income dispersion on markups paid and search intensity are most pronounced for high-income households. In Appendix Figure B10, I show that the differential effects of income dispersion on high-income households are due to market segmentation: these differences largely disappear in a single-product calibration

⁴³Formally, for each $\chi \in [0, 1]$, the counterfactual distribution $\tilde{H}(z) = H\left(\frac{z-\chi\bar{z}}{1-\chi}\right)$, where \bar{z} is average income. When $\chi = 0$, $\tilde{H}(z) = H(z)$, and in the limit $\chi \rightarrow 1$, $\tilde{H}(z)$ is a degenerate distribution with unit mass at \bar{z} .

of the model. Higher income dispersion leads to a larger increase in markups paid by high-income households because new, low search-intensity households that enter the economy buy goods that are predominantly purchased by high-income households.

Markups across cities. Finally, I use the model to predict how markups vary across CBSAs using information on the income distribution in different cities. I use data on the income distribution of each CBSA from the American Community Survey (ACS) five-year estimates. The income bins used by the ACS coincide almost one-for-one with those used by NielsenIQ, so I am able to simulate the model using the income distribution of each CBSA from the ACS with the search and spending share parameters calibrated for each income group. Note that I use the parameters $\beta_k(z)$ and $\phi(z)$ calibrated on aggregated national data to simulate markups independently for each CBSA.

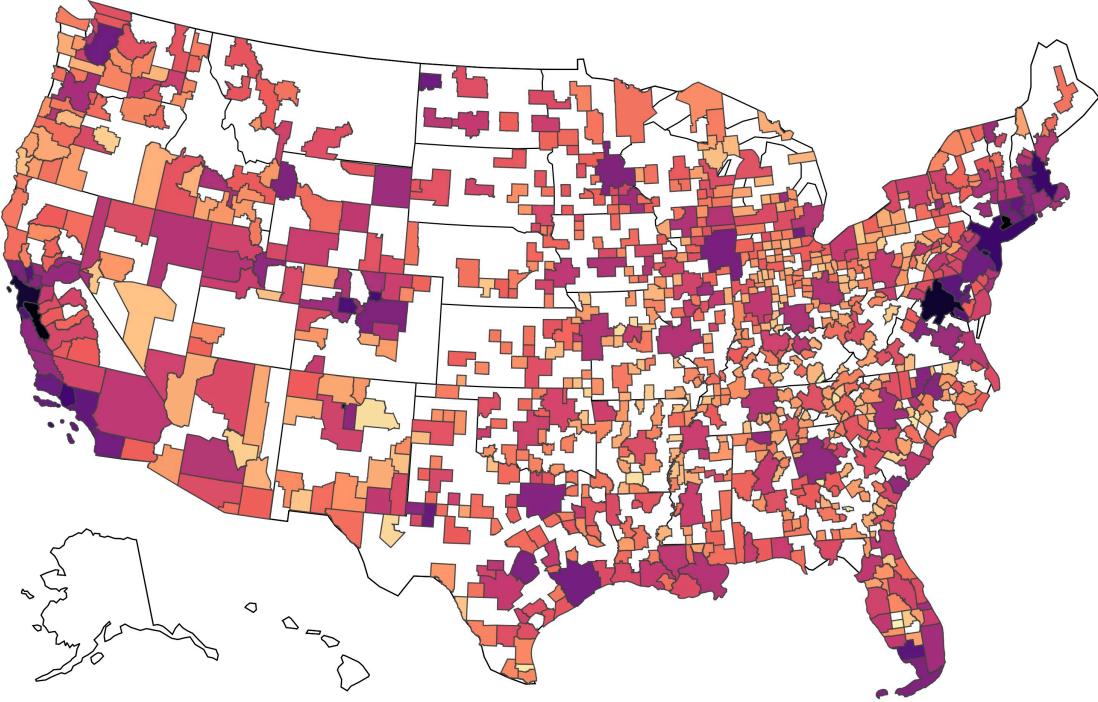
The predicted markups across CBSAs are shown in Figure 11. The model predicts particularly high markups in high-income coastal cities, such as New York, Washington, D.C., Boston, and the San Francisco Bay Area. Some less dense, high-income areas are also predicted to have high markups, such as Bridgeport, Connecticut; Napa, California; Key West, Florida; and Jackson, Wyoming. These patterns mirror the distribution of markups across CBSAs in the data (see Appendix Figure B11).

The first two columns of Table 7 show that per-capita income and income dispersion—as measured by the Gini Indices from the ACS—are important determinants of the CBSA markups predicted by the search model. Consistent with these predictions, aggregate (cost-weighted average) markups across CBSAs in the data also rise with per-capita income and inequality. Remarkably, the elasticities of CBSA markups to local income and inequality in the model and in the data are nearly identical. Appendix Table B6 shows that the relationship between CBSA markups and income inequality in the data is robust to using other measures of income inequality, such as the share of income going to the top decile reported by Sommeiller and Price (2018).

Markups predicted by the model not only replicate patterns recorded in the data, but, as shown in Table 7 column 5, account for 31 percent of the variation in CBSA markups in the data. A linear regression of markups in the data on per-capita income and inequality, for comparison, explains only 28 percent of the variation in the data.

I compare markups across CBSAs predicted by the search model with markups predicted by a representative agent, nested CES model. The nested CES model is frequently used to infer markups from data on firm market shares and concentration (see e.g., Atkinson and Burstein 2008; Smith and Ocampo 2023). Predictions from the search model focus on how demand-side factors—most importantly the income distribution—affect markups,

Figure 11: Predicted markups across CBSAs, based on CBSA income distributions.



Note: CBSAs are colored according to the aggregate markup predicted by the model, ranging from 29 percent (light yellow) to 46 percent (dark purple).

while predictions from the nested CES model focus on how supply-side factors—retailer market shares and concentration—affect markups.⁴⁴

Markups predicted by the nested CES model account for only 10 percent of the variation in CBSA markups in the data and are negatively associated with markups in the data. Column 7 shows that when income measures, predicted markups from the search model, and predicted markups from the nested CES model are all used to predict markups across CBSAs together, the markups predicted by the search model continue to be positively associated with markups in the data, while the markups predicted by the nested CES model continue to be negatively associated with markups in the data. Moreover, including income measures and markups predicted by the nested CES model does not materially increase the amount of variation explained in the data beyond using the search model-predicted markups alone.

⁴⁴To calibrate the representative agent nested CES model, I use market shares of each retailer in each CBSA for each of the K product groups. The elasticity facing retailer r for good k is $\sigma_{rk} = \rho + s_{rk}(\eta - \rho) + s_k(1 - \eta)$, where s_{rk} is retailer r 's market share in good k , s_k is the market share of good k across all K goods, ρ is the elasticity of substitution across retailers selling the same good, and η is the elasticity of substitution across the K goods. I assume retailers set markups according to the Lerner formula $\mu_{rk} = \sigma_{rk}/(\sigma_{rk} - 1)$. I choose $\rho = 5$ and $\eta = 2$ to match the level and dispersion of markups across CBSAs in the data.

Table 7: Search model-predicted markups across cities rise with per-capita income and inequality, consistent with markups across cities in the data.

Log CBSA Markup	Model-Predicted		Data				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log CBSA Income	0.102** (0.001)	0.096** (0.002)	0.110** (0.006)	0.102** (0.007)			0.018 (0.015)
Gini Index		0.113** (0.014)		0.153** (0.057)			0.067 (0.058)
Log Model-Predicted Markup				1.056** (0.053)			0.824** (0.143)
Log Nested CES Markup					-0.720** (0.072)	-0.122* (0.074)	
N	881	881	881	881	881	881	881
R ²	0.84	0.85	0.27	0.28	0.31	0.10	0.31

Note: The dependent variable in columns 1–2 is the aggregate markup for the CBSA predicted by the search model, and in columns 3–7 is the aggregate (cost-weighted average) CBSA markup in the data. Regressions weighted by total CBSA expenditures. ** is significant at 5%, * at 10%.

8 Changes in the Income Distribution from 1950–2018

How do changes in the income distribution over time affect markups? In this section, I first provide model-free estimates using moments from the data, which suggest that income growth from 1950–2018 predicts an increase in the aggregate retail markup between 6 and 23pp. I then use the calibrated search model to predict how changes in the income distribution affect markups. 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.

8.1 Model-free estimates

First, consider the case of complete price discrimination, in which case each household pays a certain markup irrespective of the incomes of other households. Using the mapping from income to markups paid in Figure 1, changes in the income distribution from 1950–2018 result in a 6.4pp increase in the average retail markup from 1950–2018.

Section 3.2, however, presents evidence of spillovers across households and estimates a macro elasticity of markups to income between 8–15 percent. Since post-tax per-capita real income grew 3.5 times from 1950 to 2018, starting at an average markup of 32 percent, a back-of-the-envelope calculation suggests a change in the aggregate markup between

13–23pp:

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

This range is quite large and does not account for rising income dispersion, which, following Proposition 2, would predict a greater change in the retail markup over time.⁴⁵ Hence, I now turn to the model.

8.2 Model estimates

In order to simulate how changes in the income distribution affect markups over time, we need to take a stand on how search productivity evolves over time. Since the empirical evidence in Section 3.2 suggests that elasticities of markups to aggregate income in the time series and cross section are nearly identical, my baseline assumption is that the relationship between search productivity and income in the time series $a(z)$ is identical to the one estimated in the cross section.⁴⁶ Thus, for income distributions from 1950 to 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.

The solid blue line in the left panel of Figure 12 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 12pp increase in the aggregate retail markup. The rise in markups is mild from 1950–1980 but accelerates significantly from 1980–2000.

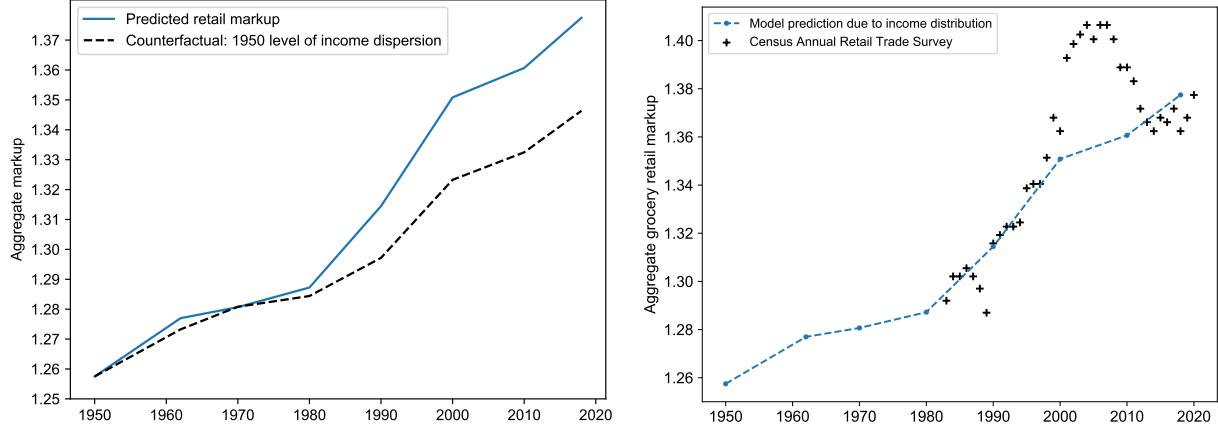
To isolate the contribution of rising income dispersion to the increase in retail markups, the dotted black line in the left panel of Figure 12 plots the predicted retail markup holding income dispersion constant at 1950 levels. The change in the predicted markup before 1980 is nearly identical to the change predicted under the actual 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 3.1pp lower than at the 2018 level of income dispersion. Table 8 summarizes the predicted change in markups and the portion due to rising income dispersion.

How does the rise in markups compare to data? The right panel of Figure 12 compares

⁴⁵ According to estimates from Saez and Zucman (2019), the share of post-tax income earned by the top 10% increased 14pp from 1950 to 2018. A back-of-the-envelope calculation using estimates from Appendix Table B6 suggests that this rise in inequality would further increase the aggregate markup $1.32 \times 0.14 \times 0.104 \approx 1.9\text{pp}$.

⁴⁶ Since $a(z)$ is increasing in z , improvements in search productivity are “baked in” to this counterfactual. In Online Appendix ??, 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 12: Predicted aggregate retail markup under income distributions from 1950–2018.



Note: In the left panel, the solid line shows the predicted aggregate retail markup from 1950 to 2018, and the dotted line shows the predicted markup holding income dispersion constant at 1950 levels. In the right panel, scatter points are markups for retail grocery stores, converted from gross margins in the Census Annual Retail Trade Survey under the assumption of constant returns.

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.⁴⁷ An aggregate markup calculated from grocery store gross margins reported by the Census increases from 29 percent in 1983 to 38 percent in 2020.⁴⁸ The path of the aggregate retail markup predicted by the model (the dashed blue line) appears to fit the secular trend in markups in the data quite well.

Notably, the data exhibit a rise and fall in markups the late 2000s not predicted by the model. Stroebel and Vavra (2019) document a link between retail markups and housing wealth that may explain this boom-bust dynamic in markups. Indeed, applying estimates of the elasticity of markups to house prices from Stroebel and Vavra (2019) appears to bridge the gap between the model predictions and the data in the late 2000s, as shown in Appendix Figure B14. Nevertheless, the long-term trend in markups is well captured by the predictions of the model.

⁴⁷Grocery stores are most analogous to the set of retailers captured in the NielsenIQ data. However, Appendix Figure B17 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.

⁴⁸Anderson et al. (2018) find that average gross margins for retail firms in Compustat rose from 0.27 to 0.31 from 1982 to 2014. Under constant returns to scale, these estimates correspond to a rise in markups from 37 to 45 percent, which is of a similar magnitude to the rise in the Census ARTS data.

Table 8: Predicted change in aggregate retail markup from 1950–2018.

Period	Predicted Δ in markup	Due to		Due to	
		Δ Income level	Δ Income dispersion	Changes in markups	Reallocations of sales
1950–2018	12.0pp	8.9pp	3.1pp	6.7pp	5.3pp
1950–1980	3.0pp	2.7pp	0.3pp	1.7pp	1.3pp
1980–2018	9.0pp	6.2pp	2.8pp	5.3pp	3.7pp

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 NielsenIQ retail scanner data is driven primarily by changes within products over time.

To calculate the contribution of reallocations to the rise in markups in the model, I assume that each quantile of the markup distribution for each good k represents a single retailer. Table 8 reports that about 60 percent of the increase in the aggregate retail markup over time in the model comes from changes in markups at each quantile of the markup distribution, while 40 percent is due to reallocations of sales across products and retailers. The reallocation of sales occurs for two reasons: preferences lead households to shift their baskets to high-markup products as incomes grow, and high-income households search less and therefore are more likely to buy from retailers with relatively high markups for each good. Appendix Figure B13 shows that these changes in search behavior leads sales of firms at the bottom of the markup distribution for a good to fall by 7 percent from 1950 to 2018, while sales of firms at the top of the markup distribution for a good grow by over 10 percent.

Effect of strategic interactions in search behavior. As aggregate incomes rise, households in the model each individually exert more search effort. These strategic interactions in search behavior moderate the change in the aggregate markup over time. If search behavior were instead held constant at 2007 levels, the model would predict a larger increase (16pp) in the aggregate markup over this period (see Appendix Table B14).

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 differ-

ences in markups paid across households arise due to utility primitives, such as differences in taste for quality and elasticities of substitution (e.g., the model developed in Online Appendix H). 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.

Robustness to calibration choices. Relaxing the assumption of Cobb-Douglas preferences across goods, using different values for the number of goods K , or allowing for pro-competitive effects do not meaningfully change the model’s predictions for the rise in markups over time. Appendix Table B13 reports results for values of K from 1 to 100, as well as with pro-competitive effects with an elasticity of search productivity to market size of $\zeta = 0.3$. In all cases, the predicted rise in markups over time varies between 10–16pp. Calibrating the model to match the within-county markup gap across income groups, rather than the unconditional markup gap, produces similar results: the predicted rise in markups over time is 11.2pp. Finally, Appendix Figure B16 shows that, even for extremely different values of σ , the predicted rise in markups over time is around 12pp, since the relative prices that households face across products do not change dramatically.

Implications for the evolution of consumption inequality. As stressed by Aguiar and Hurst (2007), differences in expenditures can be a poor proxy for differences in consumption when comparing households that have varying opportunity costs of time and search behavior. Appendix Table B16 reports the level and evolution of inequality in costs of goods purchased net of markups—as a proxy for consumption inequality—in the model. (Note that these estimates extrapolate from markups observed in the data for fast-moving consumer goods to markups paid over the entire consumption bundle, and so should be interpreted with caution.) In the baseline year, the Gini index of costs of goods purchased is 2.5 percent lower than the Gini index in post-tax income, reflecting the fact that high-income households pay higher markups. The evolution of markups in the model also suggests that inequality in costs of goods purchased has grown more slowly than inequality in post-tax income.

9 Extensions

The analysis in the main text prompts several broader questions. Here, I enumerate a few that are addressed in the Online Appendix and outline areas for future work.

When did the rise in retail markups begin? The model predicts that retail markups were rising before 1980, though the rise has accelerated in recent decades due to growing income dispersion. I compile historical estimates of retail gross margins from an NBER study by Barger (1955). The comparison of these estimates with figures from the Census of Annual Retail Trade Statistics should be interpreted with care, since methods and samples differ across these sources. Yet for several retail sectors, as shown in Appendix Figure B17, estimates from Barger (1955) suggest retail markups were lower in level and also rising far before 1980.

Markups beyond retail. While this paper's focus till now has been on retail markups, the trend of rising markups extends beyond the retail sector. To what degree can the channel proposed in this paper explain rising markups in other sectors? In some models of vertical supply chains, a decline in consumer price sensitivity can lead markups to increase along an entire chain of producers (e.g., Tirole 1988 Ch. 3).⁴⁹ In Appendix E.1.1, I provide suggestive empirical evidence for this channel, showing that De Loecker et al. (2020) markups of upstream firms are higher when they supply to retailers with high-income customer bases. These patterns are by no means conclusive, but they suggest that exploring the transmission of falling consumer price sensitivity to upstream firms merits further research.

Spatial spillovers. How do local income shocks transmit across space? The empirical evidence I present on spillovers suggests that changes in income in one area can affect households buying the same product or shopping at the same retail chain in another area. Appendix I presents a proof-of-concept that the search model developed in this paper can be used to simulate spillovers across space that arise due to uniform pricing across retail chains. Richer spatial models can incorporate these markup spillovers across space alongside other interactions.

10 Conclusion

Since Harrod (1936) first conjectured his “Law of Diminishing Elasticity of Demand,” several studies have documented a relationship between income and price sensitivity. This paper extends that literature by showing that income levels and income inequality are important determinants of markups in an economy. The relationship between markups

⁴⁹Wu (2022) suggests that the same intuitions extend to a general production network under certain conduct assumptions.

and aggregate income that I document arises both because of partial equilibrium effects—high-income households exert less search effort and buy a higher share of high-markup goods—and general equilibrium effects—firms’ markups are determined by the distribution of their customers’ incomes. Together, these effects suggest that doubling aggregate income raises markups between 8 to 15 percent.

A model that accounts for the micro evidence can help us understand a number of its macroeconomic implications, such as how changes in income level and income inequality shape prices faced by individual households across the income distribution, and why markups in the data exhibit considerable variation across cities. The model also speaks to the rise in markups observed over the past several decades. Through the lens of the model, changes in the income distribution from 1950 to 2018 can account for a substantial rise in retail markups over this period. These changes also generate a reallocation of sales to high-markup firms without any changes in firm conduct, the nature of competition, or production. This evidence suggests that markups are not a purely “supply-side” phenomenon: changes in the composition of demand may play a potent role in the evolution of markups over time.

References

- Afrouzi, H., A. Drenik, and R. Kim (2021). Growing by the masses. Working paper.
- Aguiar, M. and E. Hurst (2007). Life-cycle prices and production. *American Economic Review* 97(5), 1533–1559.
- Aguiar, M., E. Hurst, and L. Karabarbounis (2013). Time use during the great recession. *American Economic Review* 103(5), 1664–96.
- Aguirregabiria, V. (1999). The dynamics of markups and inventories in retailing firms. *The Review of Economic Studies* 66(2), 275–308.
- 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.
- Berry, S. and P. Jia (2010). Tracing the woes: An empirical analysis of the airlines industry. *American Economic Journal: Microeconomics* 2(3), 1–43.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4), 841–890.
- Bhardwaj, A., D. Ghose, S. Mukherjee, and M. Singh (2022). Million dollar plants and retail prices. Policy Research Working Paper 9995, World Bank.
- Bornstein, G. (2021). Entry and profits in an aging economy: The role of consumer inertia. Technical report, Working Paper.
- 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.
- 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.
- Caplovitz, D. (1963). *The Poor Pay More: Consumer practices of low-income families*. The Free Press.
- Chodorow-Reich, G. (2020). Regional data in macroeconomics: Some advice for practitioners. *Journal of Economic Dynamics and Control* 115(103875).
- Coibion, O., Y. Gorodnichenko, and G. H. Hong (2015). The cyclicalities 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.
- 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.
- Eichenbaum, M., N. Jaimovich, and S. Rebelo (2011). Reference prices, costs, and nominal rigidities. *American Economic Review* 101(1), 234–262.
- 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.
- Gicheva, D., J. Hastings, and S. B. Villas-Boas (2010). Investigating income effects in scanner data: Do gasoline prices affect grocery purchases. *American Economic Review: Papers & Proceedings* 100(2), 1–9.
- 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.
- Grieco, P. L., C. Murry, and A. Yurukoglu (2021). The evolution of market power in the us auto industry. Technical Report 29013, National Bureau of Economic Research.
- 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.
- Grunewald, A., J. A. Lanning, D. C. Low, and T. Salz (2020). Auto dealer loan intermediation: Consumer behavior and competitive effects. Technical Report 28136, National Bureau of Economic Research.
- 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.
- Hitsch, G. J., A. Hortaçsu, and X. Lin (2021). Prices and promotions in us retail markets: Evidence from big data. *Quantitative Marketing and Economics*, 1–80.
- Hoch, S. J., B.-D. Kim, A. L. Montgomery, and P. E. Rossi (1995). Determinants of store-level price elasticity. *Journal of Marketing Research* 32(1), 17–29.
- Hossain, M. (2020). Less is more expensive: Income differences in bulk buying. Working paper.
- Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the us retail sector. *The Quarterly Journal of Economics* 134(2), 715–783.
- Jaravel, X. (2021). Inflation inequality: Measurement, causes, and policy implications. *Annual Review of Economics* 13.
- Kaplan, G. and G. Menzio (2015). The morphology of price dispersion. *International Economic Review* 56(4), 1165–1206.
- Kaplan, G., G. Menzio, L. Rudanko, and N. Trachter (2019). Relative price dispersion. *American Economic Journal: Microeconomics* 11(3), 68–124.

- 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. (2007). Immigration and prices. *Journal of Political Economy* 115(4), 548–587.
- McKenzie, D. and E. Schargrodska (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.
- Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory* 18(2), 301–317.
- Nakamura, E. and J. Steinsson (2014). Fiscal stimulus in a monetary union: Evidence from us regions. *American Economic Review* 104(3), 753–792.
- Nakamura, E. and D. Zerom (2010). Accounting for incomplete pass-through. *Review of Economic Studies* 77, 1192–1230.
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2), 307–342.
- 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.
- Prahalad, C. K. and A. Hammond (2002). Serving the world's poor, profitably. *Harvard Business Review* 80(9), 48–59.
- 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.
- Sangani, K. (2023). Pass-through in levels and the unequal incidence of commodity shocks. Working paper.
- Seiler, S. and F. Pinna (2017). Estimating search benefits from path-tracking data: Measurement and determinants. *Marketing Science* 36(4), 565–589.
- Simonovska, I. (2015). Income differences and prices of tradables: Insights from an online retailer. *The Review of Economic Studies* 82(4), 1612–1656.
- Smith, D. A. and S. Ocampo (2023). The evolution of u.s. retail concentration. Working paper.
- Sommeiller, E. and M. Price (2018). The new gilded age: Income inequality in the us by state, metropolitan area, and county. Technical report, Economic Policy Institute.
- Stigler, G. J. (1961). The economics of information. *Journal of Political Economy* 69(3), 213–225.
- Stroebel, J. and J. Vavra (2019). House prices, local demand, and retail prices. *Journal of Political Economy* 127(3), 1391–1436.
- Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.
- Varian, H. R. (1980). A model of sales. *American Economic Review* 70(4), 651–659.

- Villas-Boas, S. B. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *Review of Economic Studies* 74(2), 625–652.
- Wolf, C. K. (2021). The missing intercept: A demand equivalence approach. Technical Report 29558, National Bureau of Economic Research.
- Wu, L. (2022). Network lerner index: Demand and distortions across industries. Working paper.

Online Appendix

(Not for publication)

A Data Cleaning and Construction	2
B Additional Tables and Figures	7
C Proofs	30
D Robustness Exercises	42
E Comparison to Other Markup Measures	50
F More Facts about Retail Markups	61
G Alternative Model: Sequential Search	71
H Alternative Model: Non-Homothetic Preferences	78
I Spatial Spillovers of Regional Income Shocks	85

Appendix A Data Cleaning and Construction

A.1 NielsenIQ Homescan

Following the NielsenIQ 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), NielsenIQ 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. These continuous measures of household income are used for estimates of the elasticity of markups to household income, for example in Appendix Table B2.

Merge with GS1. I merge UPCs from the NielsenIQ 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):

$$\hat{p}_g = \log\left(\frac{\text{Price}(g)}{\text{Units}(g)}\right) - \log\left(\frac{\sum_{g' \in \text{Module}(g)} \text{Price}(g')}{\sum_{g' \in \text{Module}(g)} \text{Units}(g')}\right). \quad (13)$$

In the above expression, $\text{Module}(g)$ is the subset of products in the same product module as g measured in the same units as g , $\text{Price}(g)$ is the price paid in transaction g , and $\text{Units}(g)$ 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. NielsenIQ 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 NielsenIQ.⁵⁰ 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 NielsenIQ Homescan data. All retail chains that are not uniquely identified in the NielsenIQ 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 NielsenIQ 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.⁵¹

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

⁵⁰Large retail chains, even those that are not part of NielsenIQ Retail Scanner program, are given unique retailer IDs.

⁵¹Stroebel 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.

principled approach than taking an unweighted average across items.

Merge with NielsenIQ 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 (13)) 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 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 NielsenIQ Homescan panel to PromoData Price-Trak market areas. Of 63,350 panelists in the 2007 NielsenIQ 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 NielsenIQ IDs). The analyses replicated using this subset of the data are reported in Table 1 and in Appendix Table B1.

Table A1: Uniformity of wholesale prices across markets.

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

Table A2: Summary statistics for markup distribution.

	<i>Measure of wholesale cost</i>	
	Base Price	Deal Price
<i>Percentiles of distribution:</i>		
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
<i>Percent below $\mu = 1$:</i>		
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.

Income group	Percent matched to wholesale cost data		Average relative unit price (\hat{p})	
	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

A.3 Other Data Sources

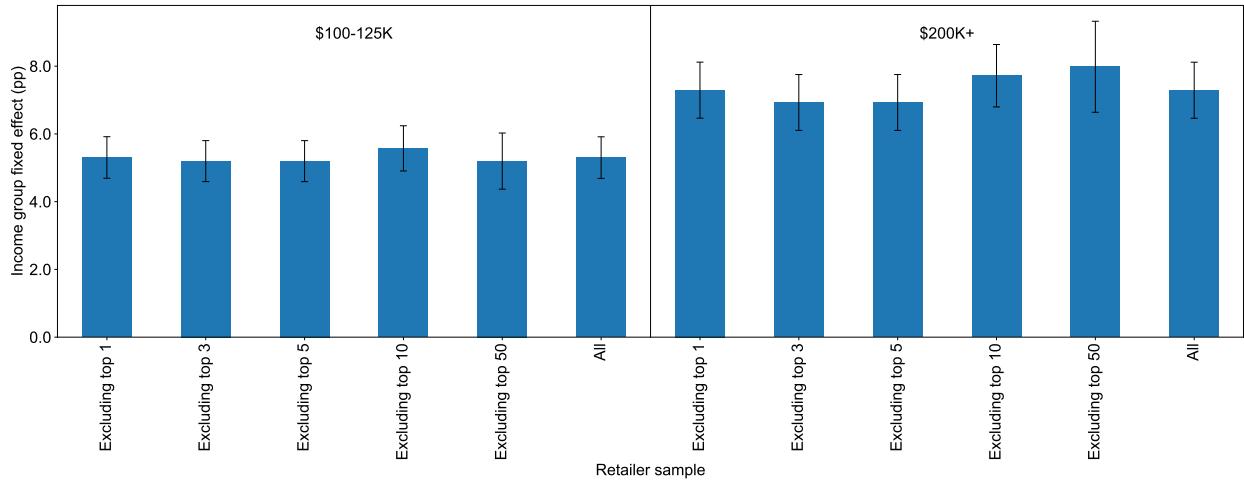
BEA estimates of county and CBSA income. Annual estimates of per-capita income by CBSA and county from 2000 to 2019 are from the BEA CAINC1 series. These income statistics use metropolitan areas delineated by the Office of Management and Budget (OMB) in bulletin 20-01, which I use to map county FIPS codes to CBSAs throughout.

CBSA income distribution and inequality measures. For CBSA income distributions, I use the number of households by income bucket from the 2009–2013 American Community Survey 5-year estimates (variable BE19001). This is the closest set of five-year estimates to 2007 that covers the vast majority of CBSAs defined in OMB bulletin 20-01. I also use Gini Indices for each CBSA from the 2009–2013 ACS 5-year estimates (variable B19083). For alternative measures of inequality, I use the shares of income going to the top 10, 5, and 1 percent of households in each CBSA from Sommeiller and Price (2018).

Grocery establishments. I use annual data on retail grocery establishments from the Census County Business Patterns. I take the number of NAICS 445 establishments (includes grocery stores, supermarkets, liquor stores, and specialty food stores) as the measure of retail grocery establishments in each county.

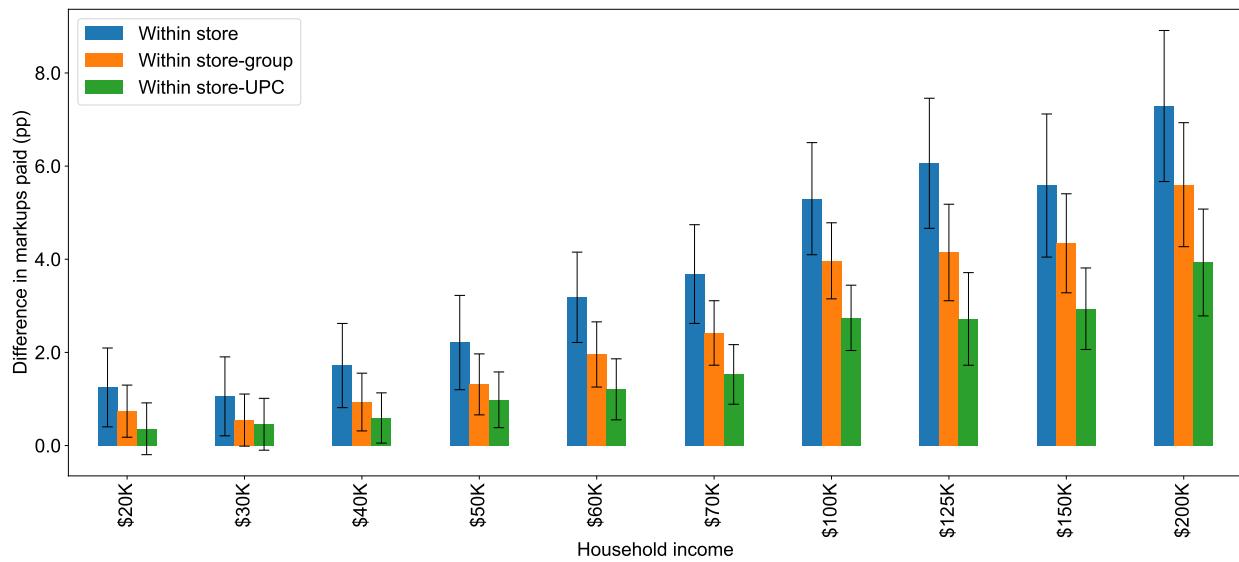
Appendix B Additional Tables and Figures

Figure B1: Stability of markup gap excluding largest retail chains.



Note: The figure shows the fixed effect estimated in specification (2) for two high-income groups as large retailers are successively removed from the sample. Retailers are ranked by total sales in the Homescan data.

Figure B2: Difference in markups paid within store (blue), within store-product group (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 cost-weighted regression of markup paid on household income dummies, 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 group fixed effects, and the green bars add store-UPC fixed effects:

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

Income levels on the horizontal axis are the minimum of the income bracket provided by NielsenIQ. Standard errors are two-way clustered by product brand and household county.

Figure B3: Savings technologies: Coupons and sales.

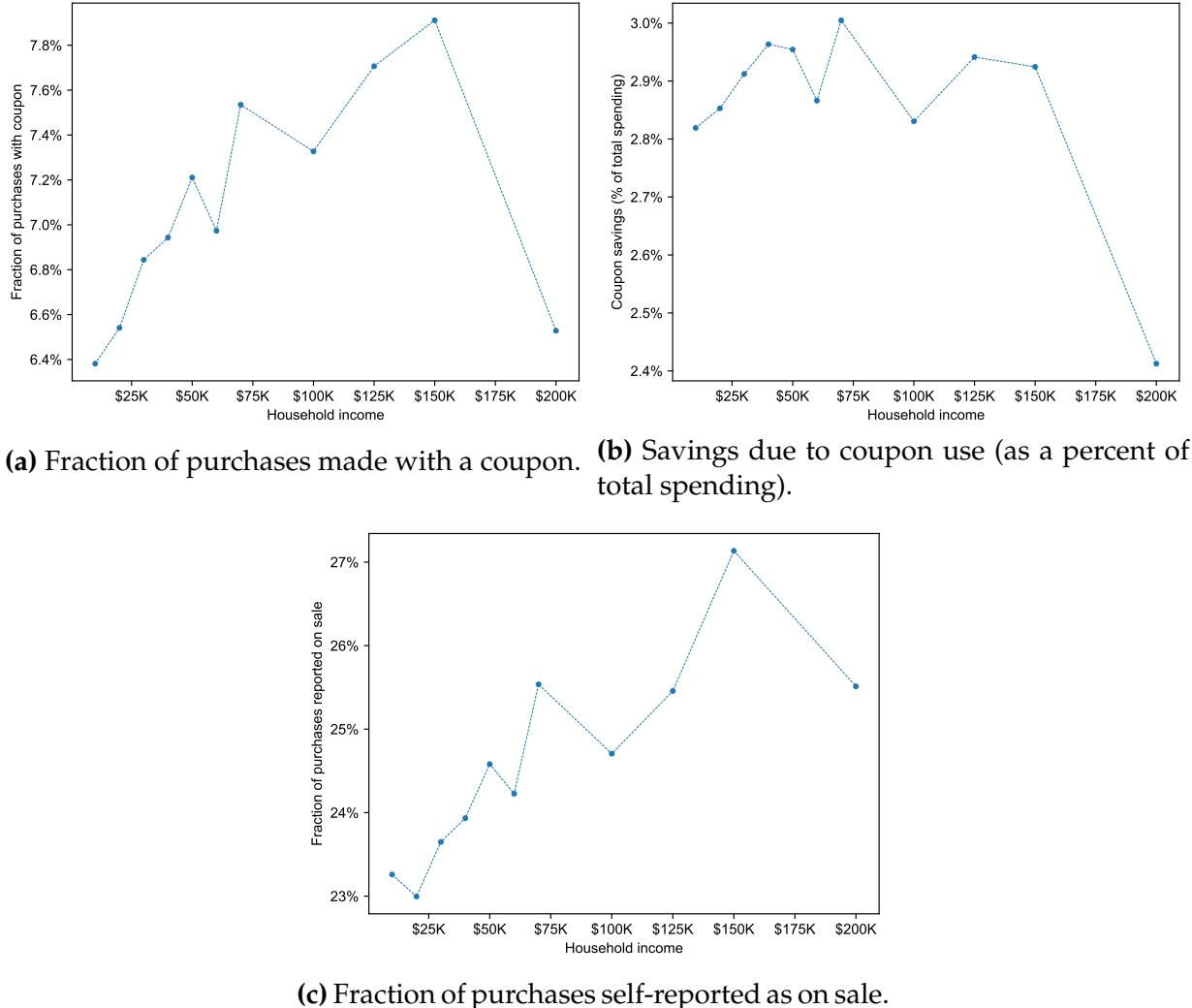
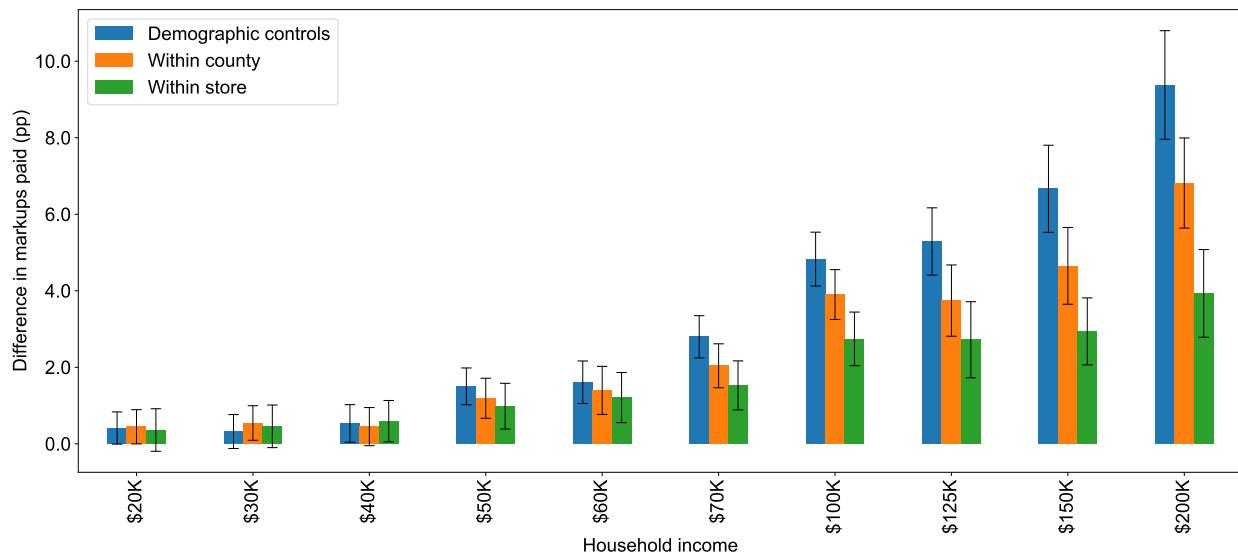
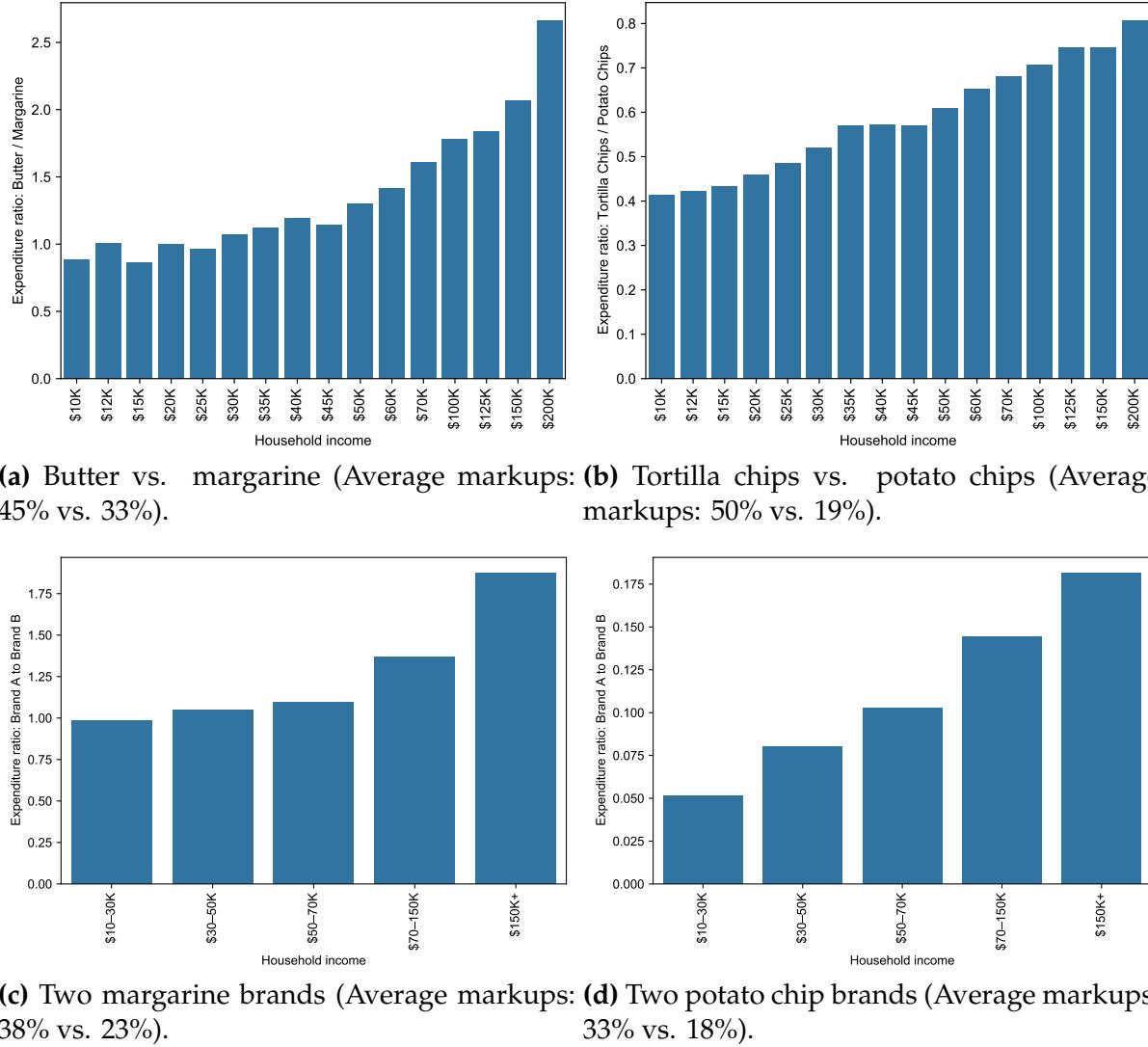


Figure B4: Difference in markups paid for identical products, relative to households with below \$20K income.



Note: The blue bars plot the coefficients β_ℓ from a cost-weighted regression of markup paid on household income dummies, demographic controls, and UPC fixed effects. The orange bars plot coefficients from an analogous regression with county-UPC fixed effects, and the green bars plot coefficients from an analogous regression with store-UPC fixed effects. Standard errors are two-way clustered by product brand and household county.

Figure B5: Examples of substitution across product modules and brands.



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 A is 38% compared to 23% for margarine brand B, and the average markup of potato chip brand A is 33% compared to 18% of potato chip brand B.

Figure B6: Household density $dH(z)$, constructed from Saez and Zucman (2019).

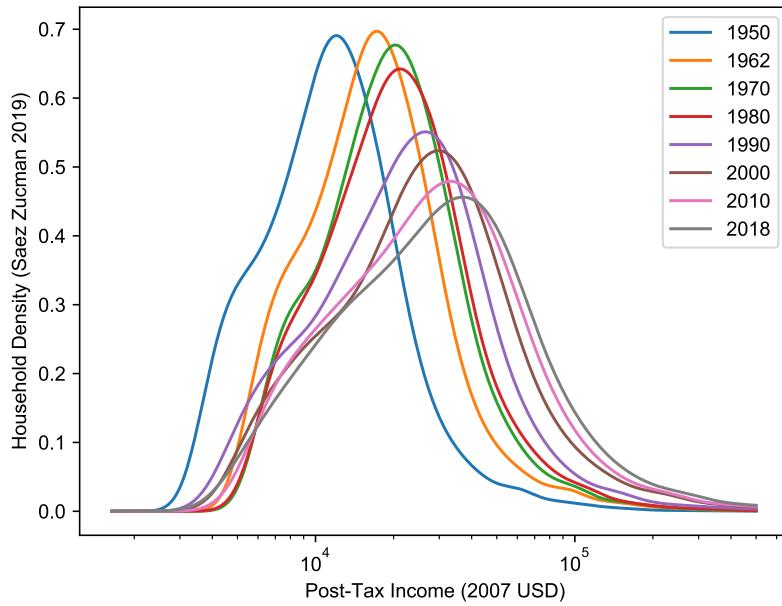


Figure B7: Segmentation of UPCs by household income.

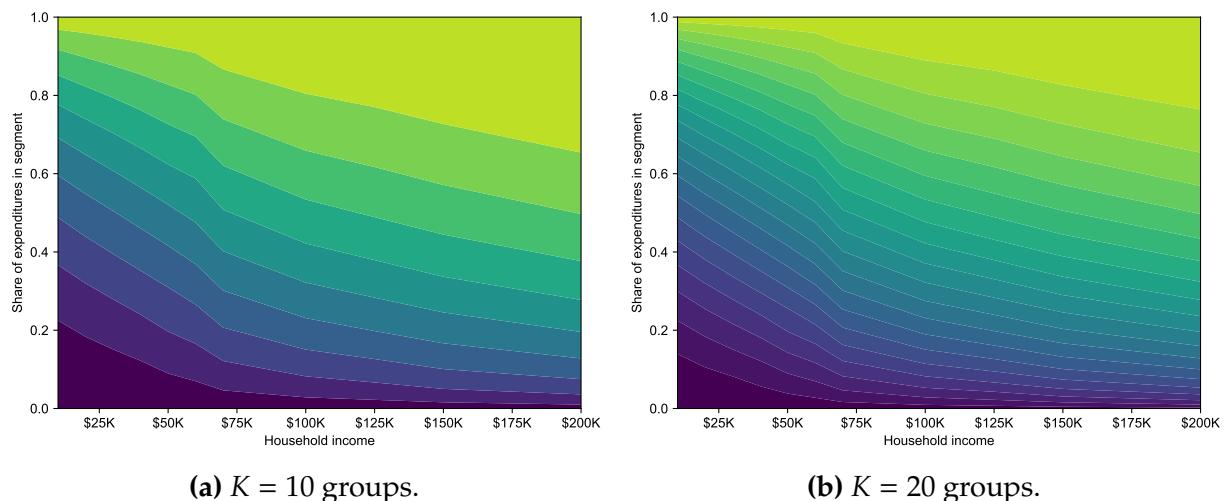
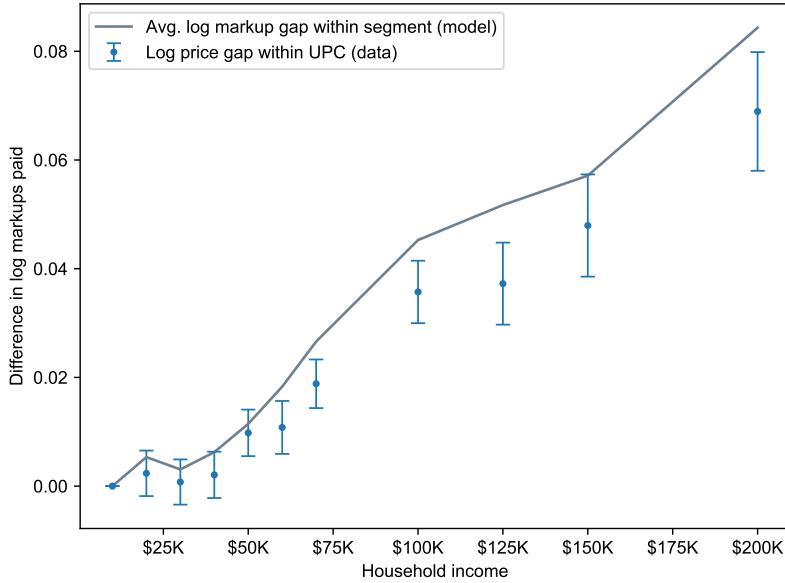


Figure B8: Difference in log markups paid for identical products in the model, compared to estimates of difference in log prices paid for identical products in the data.



Note: The grey line shows the average difference in log markups paid by each income group relative to households with below \$20K in income, in the model calibrated with $K = 10$ groups. The scatter points show coefficients from a regression of log unit prices on household income dummies, demographic controls, and UPC fixed effects.

Figure B9: Comparison of price sensitivity in calibrated model to estimates of elasticity of substitution from Auer et al. (2022).

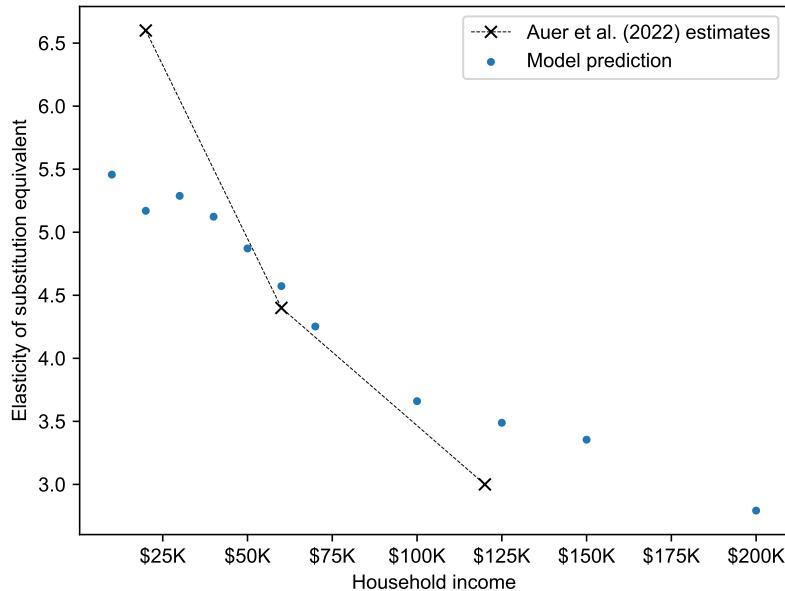


Figure B10: Effects of income inequality on markups in baseline and single-good models.

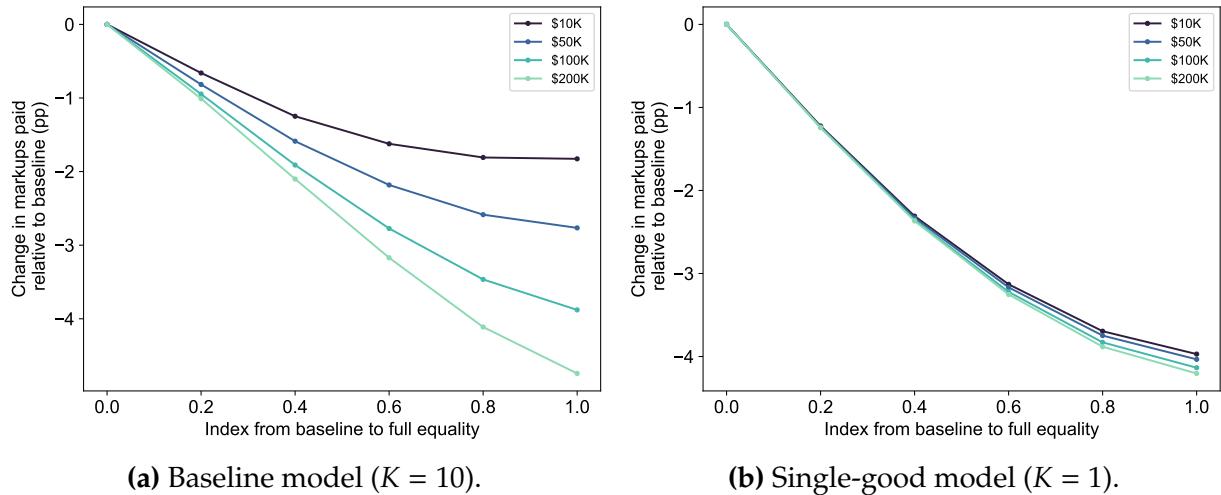
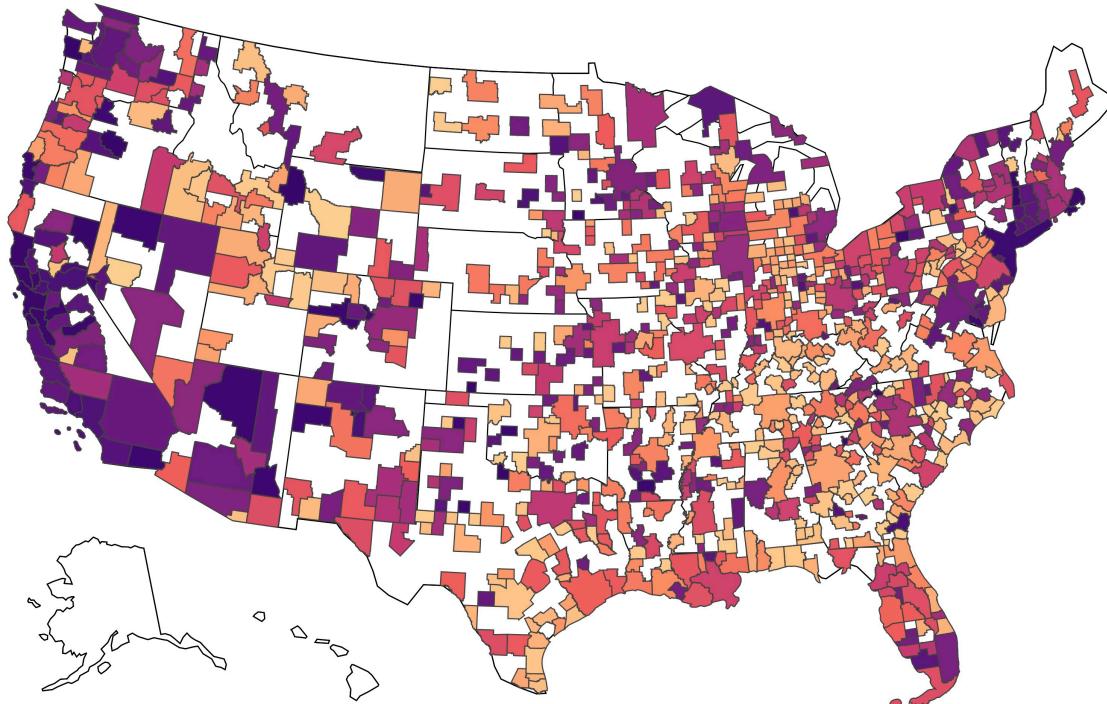


Figure B11: Markups across CBSAs in the data.



Note: CBSAs are colored according to expenditure-weighted percentile of aggregate markups across CBSAs in the data. Colors range from a percentile of zero (light yellow, corresponding to an aggregate markup of 0.95) to percentile of 1 (dark purple, corresponding to an aggregate markup of 1.70).

Figure B12: Predicted search intensity by real income level (2007 USD) from 1950–2018.

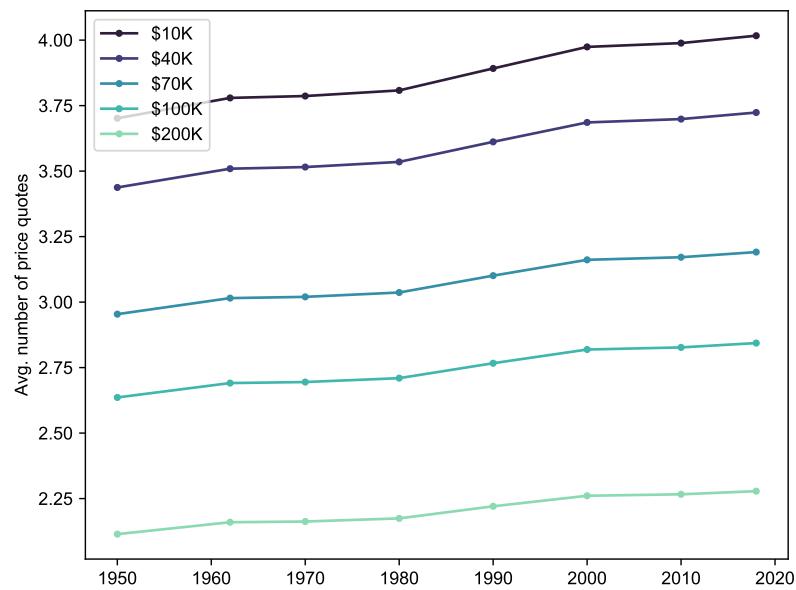
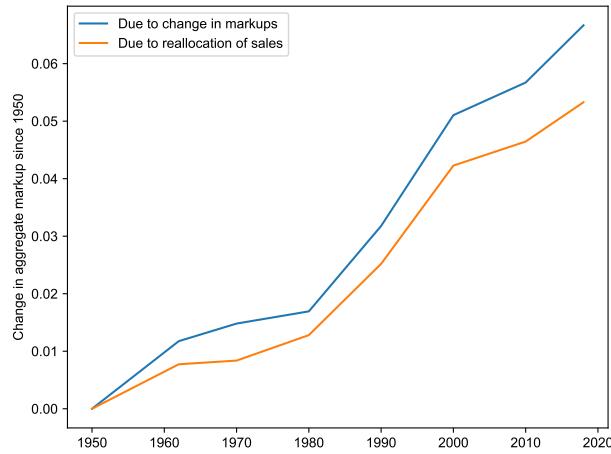
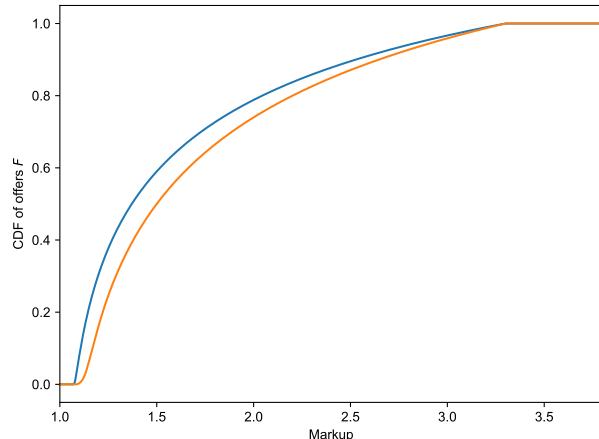


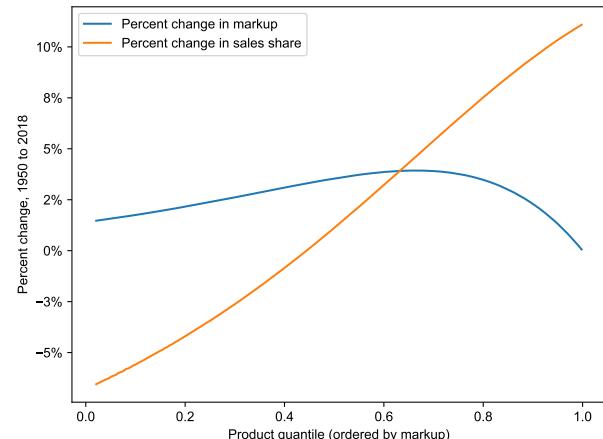
Figure B13: Role of reallocations and change in the offer price distribution, 1950–2018.



(a) Decomposition of change in agg. markup.

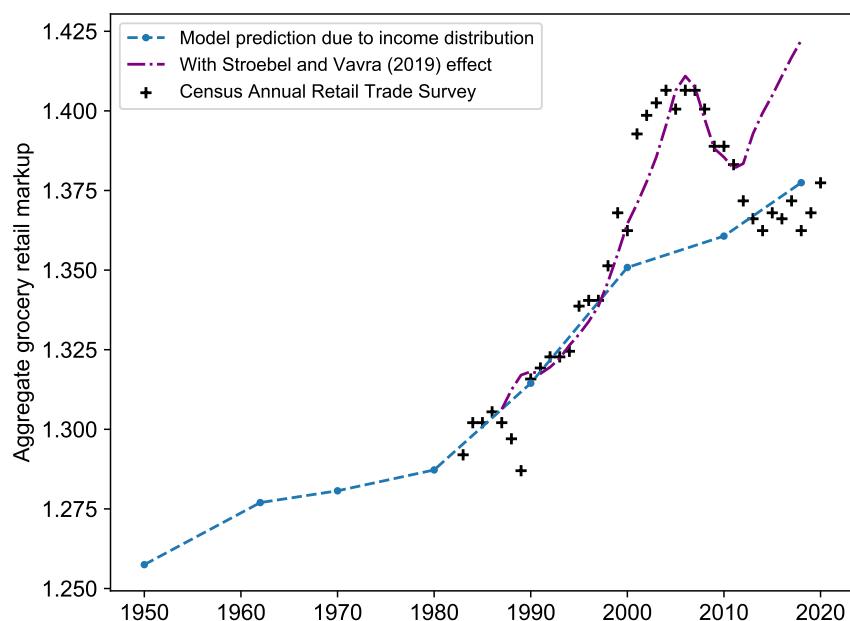


(b) Offer distribution F in 1950 and 2018.



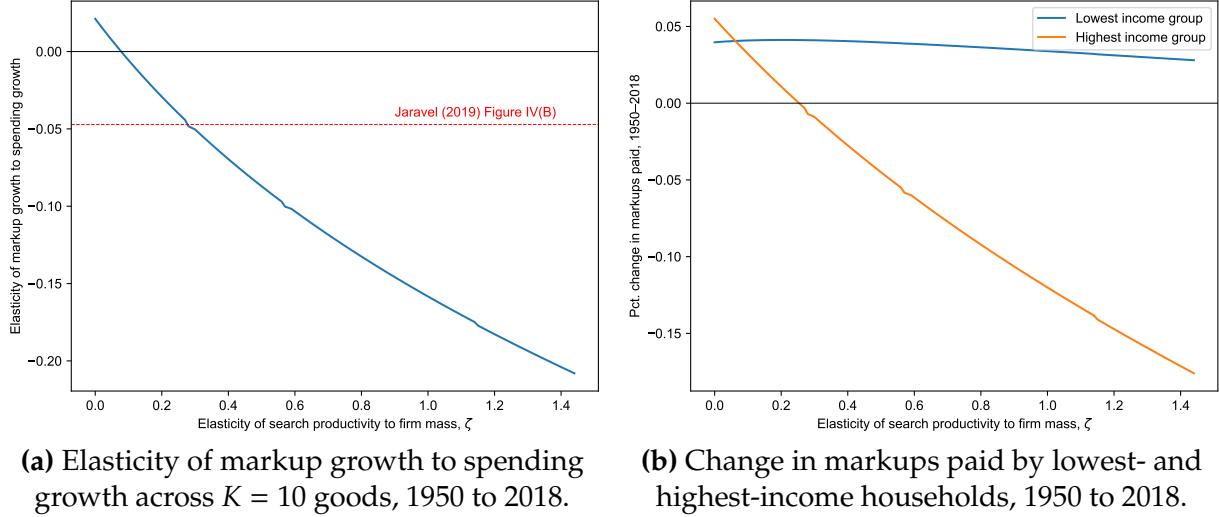
(c) Percent change in markups and sales shares by quantile for $k = 1$, 1950 to 2018.

Figure B14: Predicted aggregate retail markup from 1950–2018, adding Stroebel and Vavra (2019) effect of housing wealth.



Note: Scatter points are markups for retail grocery stores, converted from gross margins in the Census Annual Retail Trade Survey under the assumption of constant returns. The dashed blue line plots markups predicted by the model, and the dash-dotted purple line adds the effect of housing wealth on retail markups. I use an elasticity of retail markups to house prices of 7 percent, which is the midpoint of the 6–8 percent range of OLS estimates from Stroebel and Vavra (2019), to changes in the S&P / Case-Shiller national home price index since 1987.

Figure B15: Calibrating the pro-competitive effect parameter, ζ .



(a) Elasticity of markup growth to spending growth across $K = 10$ goods, 1950 to 2018.

(b) Change in markups paid by lowest- and highest-income households, 1950 to 2018.

Note: The left panel plots the coefficient estimated on a regression of markup growth on real spending growth across $K = 10$ products, simulating the model under the income distributions from 1950 and 2018 with various values of ζ . The red dotted line corresponds to the OLS estimate of -0.047 from Jaravel (2019) Figure IV(B). The right panel plots changes in markups paid by households with \$10K and \$200K in income (2007 USD) from 1950 to 2018 under various values of ζ .

Figure B16: Robustness to choice of σ : Model predicted spending shares in 1950 and predicted change in markups over time.

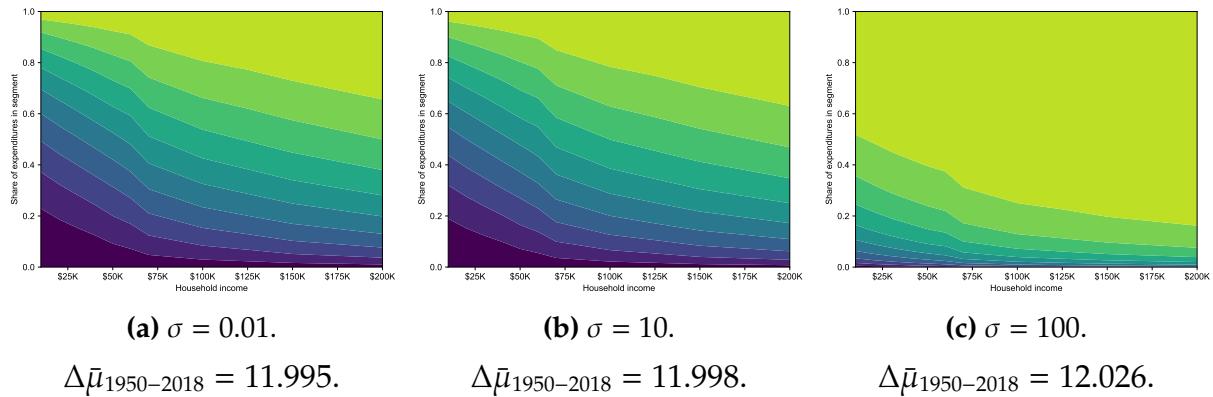
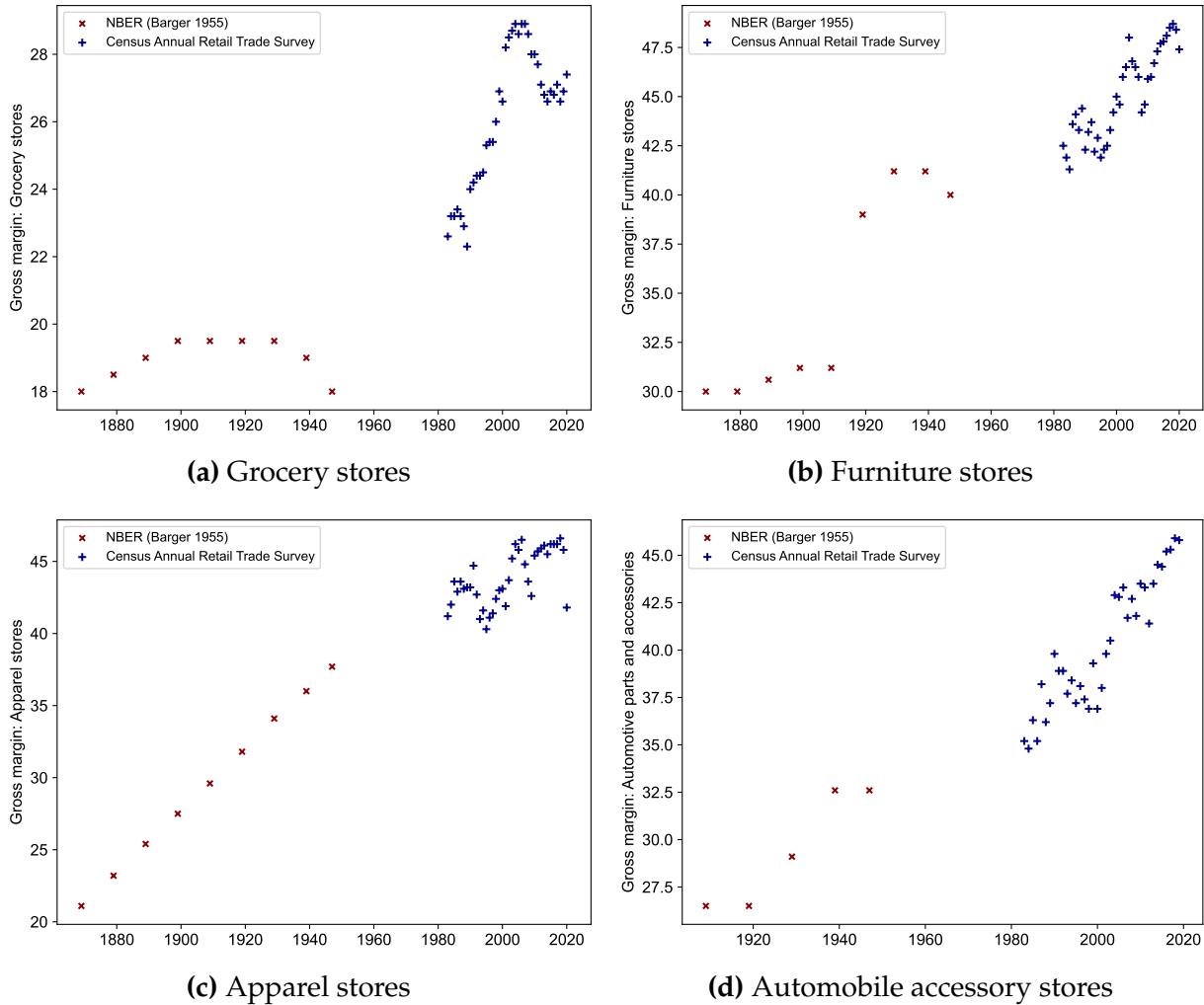


Figure B17: Data on retail gross margins over time by subsector.



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.

Income group	Share of markup gap within county (%)		Share of markup gap within store (%)	
	\$100–125K	\$200K+	\$100–125K	\$200K+
Baseline (2007 data)	81.2	78.6	65.4	54.6
Log markups	84.0	81.4	66.8	57.5
Weighting by sales	74.3	71.5	60.9	47.7
Using PromoData base price	82.5	79.5	69.4	59.2
Using PromoData market-level price	77.9	71.7	60.6	52.9
Without demographic controls	83.7	81.2	71.5	57.2
With day-of-week fixed effects	81.2	78.7	64.9	54.0
Excluding perishable items	80.3	78.1	64.7	54.3
2006 data	80.7	79.9	66.7	56.1
2008 data	83.6	78.5	68.1	57.1
2009 data	87.1	81.7	65.6	58.6

Table B2: Impact of own income on markups paid and prices paid for identical products.

	Log Markup					Log Unit Price	
	Full Sample			Store ID Sample		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Household Income	0.032** (0.004)	0.038** (0.005)	0.029** (0.004)	0.021** (0.002)	0.017** (0.002)	0.017** (0.001)	0.020** (0.001)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	
County FEs		Yes		Yes	Yes		
Store FEs				Yes			
UPC FEs					Yes	Yes	
N (millions)	25.8	25.8	25.8	14.0	14.0	25.8	25.8
R ²	0.00	0.00	0.02	0.03	0.09	0.96	0.96

Note: Demographic controls include race, ethnicity, household size, 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 B3: Impact of own income on markups paid and prices paid for identical products, instrumenting for household income.

	Log Markup					Log Unit Price	
	Full Sample			Store ID Sample		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Household Income	0.037** (0.007)	0.057** (0.008)	0.051** (0.008)	0.040** (0.005)	0.031** (0.004)	0.016** (0.002)	0.024** (0.002)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	
County FEs		Yes	Yes	Yes	Yes		
Store FEs				Yes	Yes		
UPC FEs						Yes	Yes
N (millions)	25.8	25.8	25.8	14.0	14.0	25.8	25.8
R ²	0.00	0.00	0.02	0.02	0.08	0.96	0.96

Note: Log household income is instrumented using the education and occupation of the male and female heads of household. Demographic controls include race, ethnicity, household size, 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 own and other buyers' incomes on markups paid in the cross section.

<i>Log Retail Markup</i>	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Log Household Income	0.032** (0.004)	0.054** (0.010)	0.021** (0.003)	0.038** (0.007)	0.013** (0.001)	0.022** (0.003)
Log CBSA Income	0.102** (0.011)	0.089** (0.011)				
Log Income at Retailer Locations			0.140** (0.039)	0.133** (0.037)		
Log Income of Other UPC Buyers					0.146** (0.043)	0.143** (0.043)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FEs		Yes	Yes	Yes	Yes	Yes
Store FEs				Yes	Yes	Yes
N (millions)	23.8	23.8	9.0	9.0	14.0	14.0
R ²	0.01	0.01	0.03	0.03	0.10	0.10

Note: In columns 2, 4, and 6, log household income is instrumented using the education and occupation of the male and female heads of household. Demographic controls include race, ethnicity, household size, 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: Robustness of spillovers exploiting variation over time from 2006–2012.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Log Retail Markup</i>										
Log Avg. CBSA Income	0.082** (0.016)	0.071** (0.013)								
Log Income at Retailer's Locations			0.073** (0.025)	0.087** (0.017)	0.069** (0.029)	0.073** (0.026)	0.073** (0.027)	0.065** (0.027)		
Log Income of Other UPC Buyers								0.154** (0.041)	0.145** (0.038)	0.142** (0.038)
Household-Income Level FEs	Yes									
Year FEs	Yes	Yes	Yes	Yes	Yes			Yes	Yes	
Household-Year FEs										Yes
County FEs	Yes		Yes							
Store FEs		Yes		Yes		Yes		Yes		
County-Year FEs										
Store-Year FEs										
N (millions)	133	91.9	50.8	50.8	50.8	50.8	50.8	97.0	97.0	97.0
R ²	0.17	0.19	0.17	0.19	0.17	0.19	0.21	0.19	0.21	0.21

Note: The sample includes all household transactions matched with wholesale cost data from 2006 to 2012. Regressions weighted by sales (in 2007 USD). Standard errors two-way clustered by brand and county. * indicates significance at 10%, ** at 5%.

Table B6: Relationship between CBSA aggregate markups, income, and inequality.

<i>Log CBSA Markup</i>	(1)	(2)	(3)	(4)	(5)
Log CBSA Income	0.110** (0.006)	0.102** (0.007)	0.092** (0.007)	0.094** (0.008)	0.100** (0.008)
Gini Index		0.153** (0.057)			
Top 10% Income Share			0.104** (0.023)		
Top 5% Income Share				0.078** (0.024)	
Top 1% Income Share					0.059** (0.029)
<i>N</i>	881	881	881	881	881
<i>R</i> ²	0.27	0.28	0.29	0.28	0.28

Note: CBSA per-capita incomes are from the BEA, Gini indices are from the American Community Survey (ACS) 5-year estimates, and top income shares are from Sommeiller and Price (2018). Regressions are weighted by CBSA sales. * indicates significance at 10%, ** at 5%.

Table B7: Robustness: Effect of own income and county income on search intensity.

Dependent variable (Measure of search intensity)	Log Own Income Coefficient	SE	Log County Income Coefficient	SE
<i>OLS estimates:</i>				
Log Shopping trips per \$1K spent	-0.26**	(0.00)	0.03	(0.02)
Log Shopping trips per transaction	-0.12**	(0.00)	0.11**	(0.03)
Log Shopping trips per brand bought	-0.12**	(0.00)	0.10**	(0.04)
Log Shopping trips per UPC bought	-0.12**	(0.00)	0.10**	(0.04)
Log Unique stores visited per \$1K spent	-0.22**	(0.00)	0.20**	(0.06)
Log Unique stores visited per transaction	-0.07**	(0.00)	0.29**	(0.06)
Log Unique stores visited per brand bought	-0.07**	(0.00)	0.27**	(0.05)
Log Unique stores visited per UPC bought	-0.08**	(0.00)	0.28**	(0.05)
Log Unique retailers visited per \$1K spent	-0.15**	(0.00)	0.06**	(0.02)
Log Unique retailers visited per transaction	-0.01**	(0.00)	0.14**	(0.03)
Log Unique retailers visited per brand bought	-0.01**	(0.00)	0.13**	(0.03)
Log Unique retailers visited per UPC bought	-0.02**	(0.00)	0.13**	(0.03)
<i>Instrumenting for household income:</i>				
Log Shopping trips per \$1K spent	-0.40**	(0.01)	0.09**	(0.02)
Log Shopping trips per transaction	-0.30**	(0.01)	0.19**	(0.03)
Log Shopping trips per brand bought	-0.27**	(0.01)	0.16**	(0.03)
Log Shopping trips per UPC bought	-0.29**	(0.01)	0.18**	(0.03)
Log Unique stores visited per \$1K spent	-0.34**	(0.01)	0.26**	(0.07)
Log Unique stores visited per transaction	-0.24**	(0.01)	0.36**	(0.06)
Log Unique stores visited per brand bought	-0.21**	(0.01)	0.33**	(0.06)
Log Unique stores visited per UPC bought	-0.23**	(0.01)	0.34**	(0.06)
Log Unique retailers visited per \$1K spent	-0.26**	(0.01)	0.11**	(0.02)
Log Unique retailers visited per transaction	-0.15**	(0.01)	0.21**	(0.03)
Log Unique retailers visited per brand bought	-0.12**	(0.01)	0.18**	(0.02)
Log Unique retailers visited per UPC bought	-0.14**	(0.01)	0.19**	(0.03)

Note: The table reports coefficients β and γ estimated from a regression of various measures of search intensity (for household i located in county c) on own income and average county income, controlling for the number of retail establishments in the county and state fixed effects:

$$\text{SearchIntensity}_{i,c} = \beta \text{Log Income}_i + \gamma \text{Log County Income}_c + \delta \text{Log Grocery Estabs}_c + \kappa_{\text{State}(c)} + \varepsilon_{i,c}.$$

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.) In the second half of the table, household income is instrumented for with the education and occupation of the male and female heads of household. Standard errors clustered by household county. * indicates significance at 10%, ** at 5%.

Table B8: Search time increases with basket size: Cross-sectional evidence.

	Log Shopping Trips		Log Unique Stores		Log Unique Retailers	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
Log Expenditures	0.588** (0.004)	0.617** (0.013)	0.176** (0.004)	0.044** (0.012)	0.432** (0.004)	0.159** (0.013)
Log Avg. Markup Paid	-0.688** (0.025)	-0.694** (0.025)	-0.172** (0.028)	-0.144** (0.028)	-0.344** (0.025)	-0.285** (0.025)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Income Level FEs	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	63 314	63 314	63 314	63 314	63 314	63 314
R ²	0.39	0.39	0.38	0.36	0.33	0.25

Note: Expenditures are total household expenditures in the NielsenIQ data, and avg. markup paid is the aggregate markup paid by the household over all observed purchases. Demographic controls include race, ethnicity, and presence and age group of female head of household. In columns 2, 4, and 6, log household size is used as an instrument for log expenditures. Standard errors clustered by household county. * indicates significance at 10%, ** at 5%.

Table B9: Search time increases with basket size: Time series evidence.

	Log Shopping Trips		Log Unique Stores		Log Unique Retailers	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
Log Expenditures	0.745** (0.010)	0.143** (0.029)	0.527** (0.006)	0.203** (0.032)	0.193** (0.002)	0.115** (0.029)
Household FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	917 692	917 692	917 692	917 692	917 692	917 692
R ²	0.92	0.87	0.87	0.85	0.77	0.77

Note: Expenditures are total household expenditures in the NielsenIQ data. In columns 2, 4, and 6, log household income is used as an instrument for log expenditures. Standard errors two-way clustered by household and year. * indicates significance at 10%, ** at 5%.

Table B10: Gap in prices paid for identical products declines with ticket size.

<i>Log Price</i>	(1)	(2)	(3)	(4)	(5)
Log Household Income	0.017** (0.003)	0.019** (0.003)	0.017** (0.003)	0.018** (0.003)	0.015** (0.002)
Log Household Income × Log Ticket Size	-0.004** (0.001)	-0.004** (0.001)	-0.004** (0.001)	-0.003** (0.001)	-0.002** (0.001)
UPC FEs	Yes	Yes	Yes	Yes	Yes
Demographic Controls		Yes	Yes	Yes	Yes
County FEs			Yes	Yes	Yes
Store FEs					Yes
N (millions)	59.8	59.8	59.8	29.6	29.6
<i>R</i> ²	0.99	0.99	0.99	0.98	0.98

Note: Demographic controls include race, ethnicity, household size, presence of female head of household, and age group of female head of household. Ticket size is the average expenditures on a UPC over all shopping trips that include that UPC. Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

Table B11: Elasticity of markups paid to CBSA income is larger for high-income households, both in cross-section and time series.

<i>Log Retail Markup</i>	2007	All years, 2006–2012	
	(1)	(2)	(3)
Log Avg. CBSA Income	0.089** (0.011)	0.073** (0.018)	0.065** (0.016)
Log Avg. CBSA Income × Mid-Income Group	0.016* (0.008)	0.021** (0.008)	0.006 (0.005)
Log Avg. CBSA Income × High-Income Group	0.030** (0.013)	0.034** (0.013)	0.010 (0.008)
Year FEs	Yes	Yes	Yes
Demographic controls	Yes		
Household FEs		Yes	Yes
County FEs		Yes	Yes
Store FEs			Yes
<i>N</i> (millions)	23.8	133	92
<i>R</i> ²	0.01	0.16	0.18

Note: Demographic controls include race, ethnicity, presence of female head of household, and age group of female head of household. Mid-Income Group is an indicator equal to one for households with incomes between \$50–100K, and High-Income Group is an indicator equal to one for households with incomes over \$100K. Regression weighted by sales (in 2007 USD), and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

Table B12: Comparison of markup distribution in data to model.

Percentile of markup distribution	\$20–\$30K		\$50–\$60K		\$100–\$125K		Over \$200K	
	Data	Model	Data	Model	Data	Model	Data	Model
10	0.78	1.11	0.82	1.12	0.89	1.13	0.94	1.15
25	1.02	1.14	1.04	1.15	1.08	1.17	1.13	1.19
50	1.25	1.19	1.26	1.20	1.31	1.24	1.37	1.28
75	1.50	1.32	1.52	1.34	1.58	1.42	1.65	1.51
90	1.80	1.58	1.81	1.62	1.89	1.78	1.99	1.98

Table B13: Calibration results varying number of product groups (K) and pro-competitive parameter (ζ).

No. groups (K)	Predicted Δ markup 1950–2018		Markup externality low-income		Markup externality high-income	
	$\zeta = 0$	$\zeta = 0.3$	$\zeta = 0$	$\zeta = 0.3$	$\zeta = 0$	$\zeta = 0.3$
1	15.7pp	12.8pp	+10pp	+8pp	-19pp	-11pp
3	12.6pp	10.9pp	+7pp	+6pp	-13pp	-6pp
5	12.2pp	10.7pp	+6pp	+6pp	-12pp	-6pp
10	12.0pp	10.6pp	+6pp	+6pp	-12pp	-5pp
20	11.9pp	10.6pp	+6pp	+6pp	-11pp	-5pp
50	11.8pp	10.5pp	+6pp	+6pp	-11pp	-5pp
100	11.8pp	10.5pp	+6pp	+6pp	-11pp	-5pp

Table B14: 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	Δ Income dispersion
1950–2018	15.5pp	11.3pp	4.2pp
1950–1980	3.8pp	3.5pp	0.4pp
1980–2018	11.6pp	7.8pp	3.9pp

Table B15: Predicted change in aggregate retail markup from 1950–2018 under different values of reservation price R .

R	Predicted Δ in markup	Due to		Due to	
		Δ Income level	Δ Income dispersion	Change in markups	Reallocations of sales
3.0	12.3pp	9.0pp	3.3pp	6.9pp	5.4pp
3.3 (Baseline)	12.0pp	8.9pp	3.1pp	6.7pp	5.3pp
4.0	11.5pp	8.6pp	2.9pp	6.2pp	5.3pp

Table B16: Evolution of inequality in costs of goods purchased in the model over time.

<i>Gini indices</i>	Baseline	1950	2018	Change	
Post-tax income	46.6	–	34.0	48.7	+14.7
Costs of goods purchased (expenditures net of markups)					
Baseline ($\zeta = 0$)	45.5	-2.5%	33.6	47.5	+14.0
Pro-competitive effect ($\zeta = 0.3$)	45.5	-2.5%	33.6	47.5	+13.9

Appendix C Proofs

This appendix includes derivations for the model and proofs for all propositions in the main text.

C.1 Households

Let $p_{ik}(s_{ik}, F_k)$ denote the average price paid by household i with search intensity s_{ik} when buying good k with a distribution of prices F_k . For convenience, I drop the i and k subscripts for proofs except where necessary.

Note that the price function $p(s, F)$ can be written as

$$p(s, F) = \sum_{n=1}^{\infty} q_n(s) \mathbb{E}[p|n], \quad (14)$$

where $q_n(s)$ is the result of the search mapping $\mathcal{S} : s \mapsto \{q_n\}_{n=1}^{\infty}$ and $\mathbb{E}[p|n]$ is the expected price paid having received n independent price quotes from F . We can write

$$\mathbb{E}[p|n] = \int_{\underline{p}}^R p d(1 - (1 - F(p))^n) = \int_{\underline{p}}^R p n (1 - F(p))^{n-1} dF(p). \quad (15)$$

Integrating by parts (15) yields

$$\mathbb{E}[p|n] = \underline{p} + \int_{\underline{p}}^R (1 - F(p))^n dp. \quad (16)$$

It will be helpful to characterize the derivatives of $\mathbb{E}[p|n]$ with respect to n , which I do in Lemma 4.

Lemma 4 (Properties of the expected price upon receiving n quotes). *For any nondegenerate price distribution F , $\mathbb{E}[p|n]$ is strictly decreasing and convex in n . Thus, $\mathbb{E}[p|n] - \mathbb{E}[p|n+1] > \mathbb{E}[p|n+1] - \mathbb{E}[p|n+2]$ for any n .*

Proof. The derivatives of (16) with respect to n are

$$\begin{aligned} \frac{\partial \mathbb{E}[p|n]}{\partial n} &= \int_{\underline{p}}^R (1 - F(p))^n \log[(1 - F(p))] dp < 0, \\ \frac{\partial^2 \mathbb{E}[p|n]}{\partial n^2} &= \int_{\underline{p}}^R (1 - F(p))^n [\log[(1 - F(p))]]^2 dp > 0. \end{aligned}$$

Since $\mathbb{E}[p|n]$ is strictly decreasing and convex in n everywhere, the function $g(n) \equiv \mathbb{E}[p|n] - \mathbb{E}[p|n+1]$ is strictly positive and decreasing in n , and hence $g(n) > g(n+1)$. ■

Denote the cumulative mass function of $\{q_n\}_{n=1}^\infty$ by $\{Q_n\}_{n=0}^\infty$, with $Q_0 = 0$. Using the fact that $q_n = Q_n - Q_{n-1}$, we can rewrite the price function (14) as

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

We can use this expression to get derivatives of the price function with respect to search effort s . Denote the partial derivative of p with respect to s by p_s and the second derivative by p_{ss} . We find:

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

Lemma 5 (Properties of the price function). *For any nondegenerate price distribution F , $p_s < 0$. Additionally, under Assumption 1, $p_{ss} > 0$.*

Proof. From Lemma 4, $\mathbb{E}[p|n] - \mathbb{E}[p|n+1] > 0$, and we have assumed that $Q_n(s) \leq 0$ for all n and $Q_1(s) < 0$. Thus, $p_s > 0$. Assumption 1 requires exactly that $p_{ss} > 0$. ■

Thus, under Assumption 1, the price function p is strictly decreasing and convex in search intensity s . We will now use this in the context of the household maximization problem. Recall from the main text that the households' maximization problem is

$$\max_{l_i, \{c_{ik}, s_{ik}\}} u(\{c_{ik}\}) = \left(\sum_{k=1}^K \left(\beta_{ik} c_{ik} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad \text{s.t.} \quad \begin{cases} \sum_k c_{ik} s_{ik} / a_i + l_i = 1, \\ \sum_k p_{ik} c_{ik} = z_i l_i. \end{cases}$$

Combining the first order condition with respect to l_i with any of the K first order conditions with respect to s_{ik} yields:

$$\frac{c_{ik}}{a_i} z_i + c_{ik} \frac{\partial p_{ik}}{\partial s_{ik}} = 0.$$

Dividing through by c_{ik} and defining $\phi_i \equiv z_i/a_i$ yields (7) in the main text, reproduced here without subscripts for ease:

$$\phi = -p_s(s, F).$$

By Lemma 5, $-p_s$ is strictly decreasing, so we can invert this equation to write $s = s(\phi, F)$. It will be helpful to note:

$$\frac{\partial s}{\partial \phi} = \frac{1}{-p_{ss}} < 0, \quad \text{and} \quad \frac{\partial^2 s}{\partial \phi^2} = \frac{p_{sss}}{-(p_{ss})^3}. \quad (17)$$

Solving for consumption across each good k yields

$$\left(\frac{\beta_{ik} c_{ik}}{u_i} \right)^{\frac{-1}{\sigma}} = \frac{p_{ik} \left[1 + \frac{-\partial p_{ik}}{\partial s_{ik}} \frac{s_{ik}}{p_{ik}} \right]}{\left[\sum_{k=1}^K \left(p_{ik} \left[1 + \frac{-\partial p_{ik}}{\partial s_{ik}} \frac{s_{ik}}{p_{ik}} \right] \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}}. \quad (18)$$

This resembles the standard demand curves from a CES demand system, but with a twist: the price internalized by households when allocating expenditures now depends on the expected price paid for the product plus a premium related to the elasticity of prices to search effort.

C.1.1 Proof of Lemma 1

Proof. Using (17), we have for any k ,

$$\frac{\partial s}{\partial z} = \frac{\partial s}{\partial \phi} \frac{d\phi}{dz} = \frac{1}{-p_{ss}} \frac{d\phi}{dz}.$$

Thus, $\frac{\partial s}{\partial z} \propto -\frac{d\phi}{dz}$. Next, using $p(z) = p(s(z))$, we get

$$\frac{\partial p}{\partial z} = p_s \frac{\partial s}{\partial z} \propto \frac{d\phi}{dz},$$

using the result from Lemma 4 that $p_s < 0$. ■

C.2 Firms

Denote the total units of good k consumed by households by C_k and the mass of firms supplying good k by M_k . Define the distribution of buyers' incomes $\Lambda_k(z)$ and aggregate search behavior $\{\bar{q}_{k,n}\}_{n=1}^\infty$ as given by (8) and (9) in the main text. For convenience, I drop subscripts k except where necessary below.

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

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

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

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

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

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

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 $\bar{q}_1 > 0$. Hence, profits of all firms must be

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

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

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

Solving yields the expression for F (10) and for the minimum price \underline{p} (11).

C.2.1 Proof of Lemma 3

Proof. Define the aggregate markup for good k , $\bar{\mu}_k$, as total sales over total variable costs. We can write this as:

$$\bar{\mu}_k = \frac{\int_{\underline{p}_k}^R p D_k(p) M_k dF_k(p)}{\int_{\underline{p}_k}^R D_k(p) M_k dF_k(p)} = 1 + \frac{\int_{\underline{p}_k}^R \pi(p) M_k dF_k(p)}{C} = 1 + (R - 1) \bar{q}_{k,1}.$$

The aggregate markup in the economy is then the cost-weighted average

$$\bar{\mu} = \frac{\sum_{k=1}^K C_k \bar{\mu}_k}{\sum_{k=1}^K C_k} = 1 + (R - 1) \frac{\sum_{k=1}^K C_k \bar{q}_{k,1}}{\sum_{k=1}^K C_k}.$$

■

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

C.3 Equilibrium stability and strategic interactions

C.3.1 Equilibrium stability

It is useful to start with the case in which households are homogeneous to understand conditions on equilibrium stability. Suppose $K = 1$ and households all share a common opportunity cost of search effort ϕ . The equilibrium is described by the fixed point,

$$q_1 = q_1(s(\phi, q_1)),$$

where q_1 is the share of households receiving only one quote and s is the search intensity of all households. In response to a perturbation in ϕ , the change in q_1 is

$$dq_1 = \frac{dq_1}{ds} \left[\frac{\partial s}{\partial \phi} d\phi + \frac{\partial s}{\partial q_1} dq_1 \right].$$

Rearranging yields,

$$\frac{dq_1}{d\phi} = \frac{\frac{dq_1}{ds} \frac{\partial s}{\partial \phi}}{1 - \frac{dq_1}{ds} \frac{\partial s}{\partial q_1}}.$$

We know that $\frac{dq_1}{ds} < 0$ and $\frac{\partial s}{\partial \phi} < 0$ from (17). For this equilibrium to be stable, we must have $\frac{dq_1}{ds} \frac{\partial s}{\partial q_1} < 1$. Note that when the equilibrium is stable, the sign of $\frac{dq_1}{d\phi}$ is equal to the sign of $\frac{dq_1}{ds} \frac{\partial s}{\partial \phi}$. Thus, when $\phi(z)$ is increasing in z , $\frac{dq_1}{d\phi} > 0$.

We can extend this intuition now to heterogeneous households. The equilibrium is described by a fixed point,

$$\bar{Q}_1 = \int_I Q_1(s(\phi_i, \bar{Q})) di.$$

In response to a perturbation in any ϕ_i , we can write the change in the aggregate fraction of households receiving only one quote as

$$d\bar{Q}_1 = \underbrace{\int_I \frac{dQ_1(s_i)}{ds_i} \frac{\partial s_i}{\partial \phi_i} d\phi_i di}_{\text{Direct effect}} + \underbrace{\left(\int_I \frac{dQ_1(s_i)}{ds_i} \frac{\partial s_i}{\partial \bar{Q}_1} di \right) d\bar{Q}_1}_{\text{Indirect effect through price distribution}},$$

Note that here, we have written, with some abuse of notation, $\frac{\partial s}{\partial \bar{Q}_1}$ to indicate how s responds to a shift in \bar{Q} . Since $Q_1(s)$ is strictly decreasing in s , a change in the aggregate \bar{Q} , due to any one household changing its search intensity, must be accompanied by a change in \bar{Q}_1 .

We then solve for $d\bar{Q}_1$ as

$$d\bar{Q}_1 = \frac{\int_I \frac{dQ_1(s_i)}{ds_i} \frac{\partial s_i}{\partial \phi_i} d\phi_i di}{1 - \beta}, \quad \text{where} \quad \beta \equiv \int_I \frac{dQ_1(s_i)}{ds_i} \frac{\partial s_i}{\partial \bar{Q}_1} di. \quad (19)$$

As in the case with homogeneous households, stability requires that $\beta < 1$. Note that when $\beta < 1$, the sign of $d\bar{Q}_1$ is given by the sign of the direct effect, $\int_I \frac{dQ_1(s_i)}{ds_i} \frac{\partial s_i}{\partial \phi_i} d\phi_i di$.

C.3.2 Proof of Lemma 2

Proof. Starting with the household first order condition,

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

take the implicit derivative with respect to the change in some other household j 's search behavior s_j to get,

$$\frac{\partial s_i}{\partial s_j} = \frac{-p_{s\bar{Q}} \frac{\partial \bar{Q}}{\partial s_j}}{p_{ss}}.$$

Under Assumption 1, we have that the denominator, $p_{ss} > 0$ (Lemma 5). Note that the numerator $-p_{s\bar{Q}} \frac{\partial \bar{Q}}{\partial s_j}$ has a natural interpretation as how returns to search change when aggregate search intensity increases.

Consider the limit in which $\phi_i \rightarrow 0$ for all i . Since we have assumed \mathcal{S} is such that $\lim_{s \rightarrow \infty} Q_1(s) = 0$, in this limit we have $\bar{Q}_1 \rightarrow 0$. Under (10), in this limit $F(p)$ approaches a degenerate distribution with point mass at $p = 1$. When F is degenerate, the returns to search $-p_s = 0$.

Now suppose $\phi_i = \varepsilon > 0$ for some i , where ε is a small number close to zero. This

perturbation results in a decrease in s_i , and thus a strict increase in $\bar{Q}_1 > 0$. Following (10), the resulting $F(p)$ is no longer degenerate and the returns to search $-p_s > 0$. Thus, in the neighborhood of $\bar{Q}_1 = 0$, the derivative $-p_{s\bar{Q}} \frac{\partial \bar{Q}}{\partial s_j} < 0$. Hence, we can conclude that there exists some ϕ^{cutoff} such that if $\phi_i < \phi^{\text{cutoff}}$ for all i , then $\frac{\partial s_i}{\partial s_j} < 0$.

It remains to show that the equilibrium is stable. We can now write β from (19) as

$$\beta = \int_I \frac{dQ_1(s_i)}{ds_i} \frac{-p_{s\bar{Q}_1}}{p_{ss}} ds_i.$$

We have shown that, for $\phi_i < \phi^{\text{cutoff}}$, $\frac{-p_{s\bar{Q}_1}}{p_{ss}} > 0$, and by our assumptions on \mathcal{S} we have that $dQ_1/ds_i < 0$. Hence, $\beta < 0$, which satisfies our condition for stability $\beta < 1$. ■

C.4 Comparative statics

For all comparative statics with respect to the distribution of buyers' incomes, I consider the case where $K = 1$, and there is a single distribution of buyers' incomes. Recall from Lemma 3 that the aggregate markup when $K = 1$ is

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

Thus, the change in the aggregate markup in response to a perturbation is

$$d\bar{\mu} = (R - 1)d\bar{Q}_1.$$

Thus, to response of $\bar{\mu}$ to a perturbation in $\Lambda(z)$, we need to characterize the response of \bar{Q}_1 . Note from (19) that we can consider the direct effects of a change in the distribution of ϕ 's on \bar{Q}_1 , since equilibrium stability ensures that indirect effects do not cancel out the direct effect.

C.4.1 Proof of Proposition 1

Proof. 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 . We can write,

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

Note that $\frac{dQ_1}{ds} < 0$ for all s , and under Assumption 1, $p_{ss} > 0$. Hence, if $\frac{d\phi}{dz} > 0$, then $\frac{dQ_1}{dz} > 0$, and thus a first-order stochastic shift in $\Lambda(z)$ increases \bar{Q}_1 and $\bar{\mu}$. ■

C.4.2 Proof of Proposition 2

Proof. Since $\bar{Q}_1 = \int_0^\infty Q_1(z)d\Lambda(z)$, a mean-preserving spread in $\Lambda(z)$ increases \bar{Q}_1 if $Q_1(z)$ is increasing and convex in z . We already have from above that $Q_1(z)$ is increasing in z if ϕ is increasing in z and Assumption 1 holds. Hence, we need to now find conditions under which $Q_1(z)$ is convex in z .

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

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

Rearranging yields,

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

which is exactly the condition guaranteed by Assumption 2. Thus, if Assumptions 1 and 2 hold, and $\phi(z)$ increasing and convex in z , then Q_1 is increasing and convex in z , and a mean-preserving spread in $\Lambda(z)$ increases \bar{Q}_1 and \bar{p} . ■

C.5 Graphical Illustration

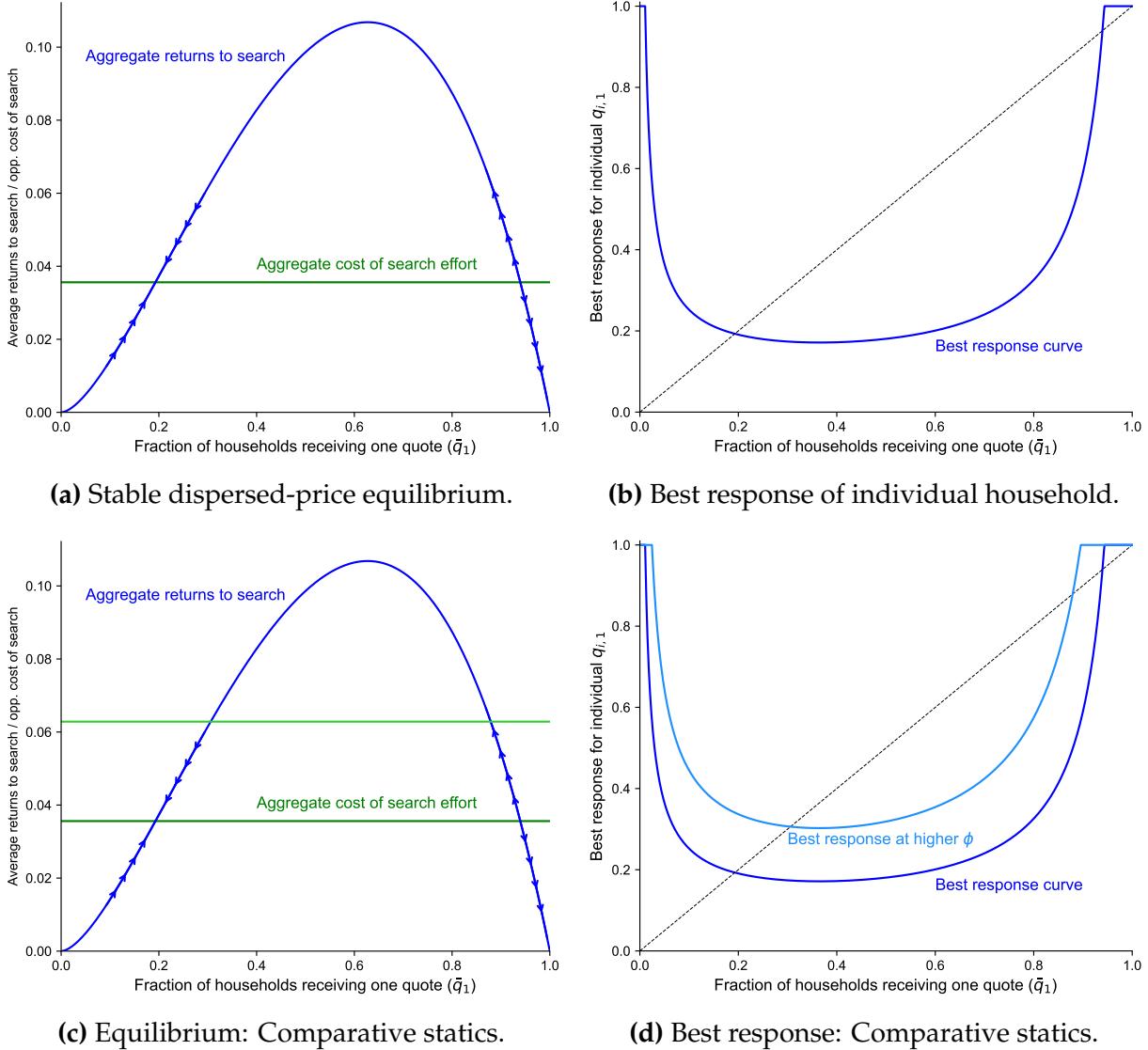
Figure C1 provides visual examples for the key model results. Aggregating households' first order conditions in (7) yields the equilibrium condition,

$$\underbrace{\int_0^\infty -p_s(s(z), \bar{Q})d\Lambda(z)}_{\text{Aggregate returns to search}} = \underbrace{\int_0^\infty \phi(z)d\Lambda(z)}_{\text{Aggregate cost of search effort}}. \quad (20)$$

Figure C1a plots these aggregate returns to search and aggregate cost of search effort as a function of the share of households receiving only one quote, \bar{q}_1 .⁵²

⁵²For plotting, I use the closed-form expressions from the two-quote parameterization in (21).

Figure C1: Stable dispersed-price equilibrium and best response curves.



Let us start at the point where $\bar{q}_1 = 0$. Since all households at this point retrieve at least two quotes, no firm can be incentivized to set its price above any other firm, and hence all firms price at marginal cost $p = 1$. Since the price distribution is degenerate, returns to search are zero for all households. At the other extreme, when $\bar{q}_1 = 1$, all households receive only quote, and hence all firms price at the reservation price $p = R$. Once again, the degenerate price distribution means that returns to search are zero.

For $\bar{q}_1 \in (0, 1)$, some households receive only one quote, the firm price distribution is no longer degenerate, and thus returns to search are strictly greater than one. There are two dispersed-price equilibria where aggregate returns to search equal aggregate cost of

search effort. The arrows on the aggregate returns to search curve denote how households update search effort when off either of equilibria. They indicate that the left-hand side equilibrium is stable, while the right-hand side equilibrium is unstable.

This is best seen in the household best response curve illustrated in Figure C1b. The best response curve plots the household's choice of search effort (proxied by $q_{i,1}$) as a function of the aggregate search effort in the economy (proxied by \bar{q}_1). For simplicity, Figure C1b shows a case with homogeneous households, so that equilibria occur where the best response curve intersects the 45 degree line, and stable equilibria occur when this intersection occurs from above. We see that the left-hand side equilibrium is the stable one. Moreover, for the case shown in Figure C1b, ϕ is small enough such that Lemma 2 holds. We can see that household search decisions are strategic substitutes by seeing that the best response curve has a negative slope when it intersects the 45 degree line at the stable equilibrium point.

Figures C1c and C1d illustrate comparative statics of the equilibrium and best response functions. When $\phi(z)$ is increasing in z , a first-order stochastic in $\Lambda(z)$ increases the aggregate cost of search effort $\int_I \phi(z) d\Lambda(z)$. As shown in Figure C1c, this shift in the aggregate cost of search increases \bar{q}_1 in the stable equilibrium and thus increases the aggregate markup (Proposition 1). A similar shift results following a mean-preserving spread in $\Lambda(z)$ when $\phi(z)$ is increasing and convex in z (Proposition 2). Finally, Figure C1d shows that an increase in ϕ for a given household increases the household's choice of $q_{i,1}$ (i.e., decreases the household's search intensity) at any value of \bar{q}_1 (Lemma 1).

C.6 Application to two-quote and Poisson cases

I show that Assumptions 1 and 2 both hold under two common parameterizations of the search mapping function S : a two-quote case and the Poisson case.

C.6.1 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

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

I show that both Assumption 1 and Assumption 2 hold for this mapping S .

Assumption 1 becomes:

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

which holds since $\exp(-s_i) > 0$ for an interior s_i and $\mathbb{E}[p|n]$ is strictly decreasing in n .

Assumption 2 becomes:

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

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

Note that closed form expressions are available for most quantities in the two-quote case:

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

Thus, the equilibrium condition (20) can be written in closed form as:

$$\underbrace{\frac{\bar{q}_1^2}{1 - \bar{q}_1} (R - mc) \left[\frac{1}{2} \frac{1}{1 - \bar{q}_1} \log\left(\frac{2 - \bar{q}_1}{\bar{q}_1}\right) - 1 \right]}_{\text{Aggregate returns to search}} = \underbrace{\int_0^\infty \phi(z) d\Lambda(z)}_{\text{Aggregate cost of time}}. \quad (21)$$

C.6.2 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,

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

I drop the i subscripts below for convenience. 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,

$$\begin{aligned}\frac{d^2Q_{n+1}}{ds^2} &= \begin{cases} e^{-s} & n = 0, \\ e^{-s} \frac{s^{n-1}}{n!} (s-n) & n \geq 1. \end{cases} \\ \frac{d^3Q_{n+1}}{ds^2} &= \begin{cases} -e^{-s} & n = 0, \\ -e^{-s} (s-2) & n = 1, \\ -e^{-s} \frac{s^{n-2}}{n!} ((s-n)^2 - n) & n \geq 2. \end{cases}\end{aligned}$$

First, we show that Assumption 1 holds:

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{d^2Q_n}{ds^2} [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] \\ &= \frac{d^2Q_1}{ds^2} [\mathbb{E}[p|1] - \mathbb{E}[p|2]] + \sum_{n=2}^{\infty} \frac{d^2Q_n}{ds^2} [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] \\ &= e^{-s} \left([\mathbb{E}[p|1] - \mathbb{E}[p|2]] + \sum_{n=2}^{\infty} \frac{s^{n-1}}{n!} (s-n) [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] \right) \\ &= e^{-s} \left(\sum_{n=1}^{\infty} \frac{s^n}{n!} [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] - \sum_{n=2}^{\infty} \frac{s^{n-1}}{(n-1)!} [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] \right) \\ &= e^{-s} \left(\sum_{n=1}^{\infty} \frac{s^n}{n!} ([\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] - [\mathbb{E}[p|n+1] - \mathbb{E}[p|n+2]]) \right) \\ &> 0.\end{aligned}$$

where the last line uses Lemma 4.

Similarly, we find that Assumption 2 holds:

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

$$\begin{aligned}
&= e^{-2s} [\mathbb{E}[p|2] - \mathbb{E}[p|3]] + \sum_{n=3}^{\infty} \left(\frac{d^2 Q_1}{ds^2} \frac{d^2 Q_n}{ds^2} - \frac{dQ_1}{ds} \frac{d^3 Q_n}{ds^3} \right) [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] \\
&= e^{-2s} [\mathbb{E}[p|2] - \mathbb{E}[p|3]] + e^{-2s} \sum_{n=3}^{\infty} \left[\frac{s^{n-1}}{(n-1)!} - \frac{s^{n-2}}{(n-2)!} \right] [\mathbb{E}[p|n] - \mathbb{E}[p|n+1]] \\
&= e^{-2s} \sum_{n=1}^{\infty} \frac{s^n}{n!} ([\mathbb{E}[p|n+1] - \mathbb{E}[p|n+2]] - [\mathbb{E}[p|n+2] - \mathbb{E}[p|n+3]]) \\
&\geq 0.
\end{aligned}$$

where the last line again uses Lemma 4.

Appendix D Robustness Exercises

D.1 Unobserved Local Costs

As in Gopinath et al. (2011), my baseline assumption when constructing retail markups is that retailers' rent and labor costs are fixed at short horizons. However, if some labor and rent costs are in fact variable costs, the presence of these unobserved local costs can bias the elasticity of markups to aggregate income measured across space. In this appendix, I gauge the potential magnitude of this bias using data on retail operating expenses and data on retail wages and rents.

To fix ideas, suppose that variable costs for a retailer with output Y are given by

$$VC(Y) = cY + wL(Y) + rA(Y),$$

where cY is the costs of goods sold, $wL(Y)$ are variable wage costs, and $rA(Y)$ are variable rent costs. I assume that production is constant returns and Leontief in merchandise, labor, and store space, so that labor $L(Y)$ and store space $A(Y)$ are linear in Y .

The aggregate markup constructed from cost of goods sold alone μ^{COGS} and the "true" aggregate markup μ^{true} are given by

$$\mu^{\text{COGS}} = \frac{pY}{cY}, \quad \text{and} \quad \mu^{\text{true}} = \frac{pY}{VC(Y)},$$

where pY are total sales.

Taking the elasticity with respect to aggregate income I , we find:

$$\frac{d \log \mu^{\text{COGS}}}{d \log I} = \underbrace{\frac{d \log \mu^{\text{true}}}{d \log I}}_{\text{True elasticity}} + \underbrace{\frac{wL}{VC} \frac{d \log w}{d \log I} + \frac{rA}{VC} \frac{d \log r}{d \log I} - \frac{wL + rA}{VC} \frac{d \log c}{d \log I}}_{\text{Bias}}. \quad (22)$$

Equation (22) shows that the bias in the measured elasticity of markups to income depends on two sets of statistics: the shares of wages and rent in total variable costs, and the elasticities of factor prices (e.g., wages and rents) to local income. If either the share of labor and rent in variable costs is zero, or if factor prices do not covary with local income, the bias in (22) disappears.

I estimate the shares of wages and rents in variable costs using data from the 2007 Census Annual Retail Trade Survey, and estimate the elasticities of wages and rents to local income using data on retail wages and rents across CBSAs. Table D1 lists costs of goods sold, labor expenses, and rent expenses for four retailer sectors using data from the 2007 Census Annual Retail Trade Survey of Detailed Operating Expenses. I intentionally choose expansive definitions that include all payroll, fringe benefits, commissions, and contractor costs for labor, and rent and repair expenses for both stores and offices / other buildings for rents. Even with these inclusive definitions, labor and rent are relatively small portions of plausibly variable costs (about 16 and 3 percent, respectively).⁵³

To estimate the elasticities of factor costs to local income, I use data on retail wages from the Bureau of Labor Statistics Occupational Employment and Wage Statistics (OEWS) and data on retail rents from Moody's REIS platform. The OEWS include mean hourly wages for cashiers and retail salespersons across 330 CBSAs in 2007. Columns 1–2 of Table D2 report that the elasticities of these hourly wages to CBSA income range from 0.26–0.29. REIS has asking and effective rents per square foot of retail space across 68 MSAs in 2007. The elasticities of these retail rents to CBSA income, reported in columns 3–4 of Table D2, are between 1.22–1.27. As discussed in the main text, both wholesale costs from the PromoData and cost data from a major grocery used by DellaVigna and Gentzkow (2019) suggest that costs of goods sold do not significantly vary with local income. For example, DellaVigna and Gentzkow (2019) report this elasticity is 0.007 (standard error: 0.005).

Figure D1 reports $d \log \mu^{\text{true}}/d \log I$ as we vary the share of labor and rent expenses we consider variable from zero to one. I use the labor and cost shares from retail grocery

⁵³Besides labor and rent expenses, other operating expenses reported in the Survey of Detailed Operating Expenses include software costs; electricity, fuel, and utility payments; purchased communication, professional and technical, and advertising and promotional services; capitalized expenses; purchases and repairs of machinery and equipment; and bad debt and interest expense.

Table D1: Retail labor and rent costs from the 2007 Census Annual Retail Trade Survey.

	All Retail	Retail Excl. Auto	Food and Beverage	Grocery stores
Sales	3,995,182	3,085,043	547,837	491,360
– Gross margin	1,105,515	941,583	160,068	141,848
Cost of goods sold	2,889,667	2,143,460	387,769	349,512
Operating expenses	873,400	736,072	128,985	116,175
Payroll	383,215	315,637	58,321	53,293
Fringe benefits	73,457	62,515	15,525	14,499
Contract labor costs	4,720	4,112	492	433
Commission expense	6,697	5,346	325	150
Total labor expenses	468,089	387,610	74,663	68,375
Lease/rental payments for stores/offices	83,134	73,435	10,262	8,525
Repairs/maintenance for stores/offices	7,257	6,231	1,385	1,281
Total rent expenses	90,391	79,666	11,647	9,806
<i>Cost shares (%):</i>				
Labor cost / (Labor cost + COGS)	13.9	15.3	16.1	16.4
Labor cost / (OPEX + COGS)	12.4	13.5	14.4	14.7
Rent cost / (Rent cost + COGS)	3.0	3.6	2.9	2.7
Rent cost / (OPEX + COGS)	2.4	2.8	2.3	2.1

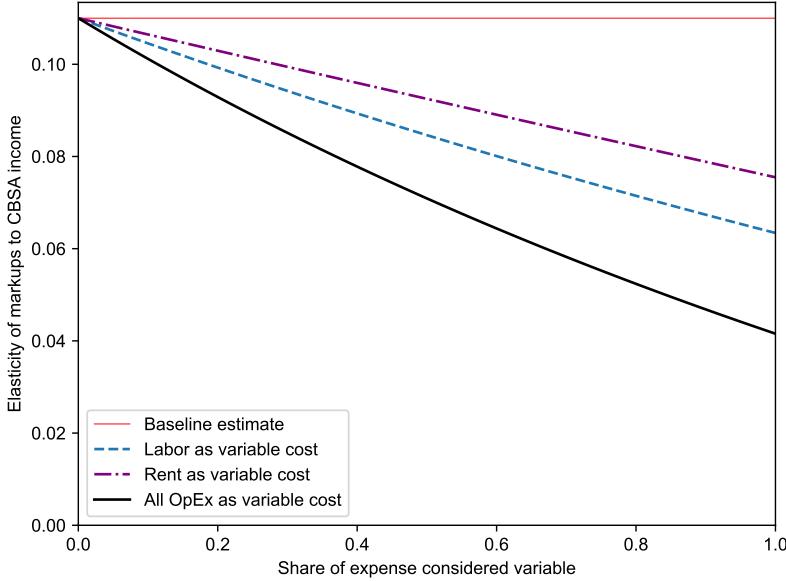
Note: Sales and gross margins are from the 2007 Census Annual Retail Trade Survey. Total operating expenses and operating expense subcategories are from the 2007 Census Retail Trade Survey of Detailed Operating Expenses.

Table D2: Elasticities of retail wages and rents to CBSA income.

Source: Variable:	Log Retail Wages		Log Retail Rents	
	OEWS Cashiers (1)	OEWS Retail Salespersons (2)	REIS Asking Rent (3)	REIS Effective Rent (4)
Log Avg. CBSA Income	0.285** (0.040)	0.265** (0.027)	1.226** (0.203)	1.265** (0.208)
N	330	330	68	68
R ²	0.19	0.23	0.42	0.43

Note: OEWS refers to the Bureau of Labor Statistics Occupational Employment and Wage Statistics. REIS refers to Moody's REIS platform. Average CBSA income from annual BEA estimates. All estimates use data from 2007. Robust standard errors in parentheses. * indicates significance at 10%, ** at 5%.

Figure D1: Robustness: Elasticity of markups to CBSA income when marginal costs include retail wages and rents.



stores in Table D1, and the larger estimates of the elasticities of wages and rents to local income from Table D2 to be conservative.

Under the baseline assumption that labor and rent costs are fixed costs, $d \log \mu^{\text{true}} / d \log I = d \log \mu^{\text{COGS}} / d \log I \approx 0.110$ (from Table 7). As we increase the share of labor and rent expenses that are considered variable, the true elasticity falls. Under the most extreme assumption that all labor and rent costs are variable, the elasticity of markups to income falls to about 4 percent. At reasonable levels, the estimated bias in the measured elasticity across CBSAs is mild. For example, if about 20 percent of either labor or rent expenses are variable, the true elasticity of markups to income is around 10 percent, and if about 20 percent of all operating expenses are variable, the true elasticity is around 9 percent.

D.2 Changes in Search Productivity Over Time

In this appendix, I test for the presence of secular improvements in search productivity over time over and above what would be expected from income growth. My test, which exploits the fact that households equate the cost of search effort with returns to search, suggests that search productivity conditional on income is flat or slightly declining over time. I also discuss what conclusions about the evolution of search intensity can be drawn from trends on shopping behavior, and compare search behavior and price dispersion over time in the model to the data.

D.2.1 Identifying secular trends in search productivity using returns to search

Recall from the household first order condition (7) (reproduced here) that households equate returns to search with opportunity cost of search effort:

$$\underbrace{-p_s(s(z), F)}_{\text{Returns to search}} = \underbrace{\phi(z)}_{\text{Opp. cost of search}} = \frac{z}{a(z)}. \quad (23)$$

I now allow search productivity to take the form $a(z, \tau)$, where τ is time and $\partial a / \partial \tau$ captures secular changes in search productivity over time. Since households equate returns to search with costs of search effort, changes in returns to search within an income group reveal changes in costs of search effort conditional on income.

Taking the derivative of (23) yields,

$$d(p_s) = \frac{1}{a} \left[- \left(1 - \frac{\partial a}{\partial z} \frac{z}{a} \right) dz + \frac{\partial a}{\partial \tau} \frac{z}{a} d\tau \right]. \quad (24)$$

If there are secular improvements in search productivity over time ($\frac{\partial a}{\partial \tau} > 0$), then returns to search $-p_s$ conditional on income should decline and hence p_s should increase.

To take this to data, I consider specification (25):

$$\begin{aligned} \text{PriceIndex}_{i,t} = & \alpha \text{SearchIntensity}_{i,t} + \beta (\text{SearchIntensity}_{i,t} \times \text{Year}_t) \\ & + \gamma (\text{SearchIntensity}_{i,t} \times \text{LogIncome}_{i,t}) + \varphi \text{Year}_t + \delta \text{LogIncome}_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (25)$$

where $\text{PriceIndex}_{i,c,t}$ is a price index for household i in year t and $\text{SearchIntensity}_{i,t}$ is a measure of i 's shopping intensity in year t .

Our coefficient of interest is β , which indicates how returns to search (conditional on own income) change over time. Mapping to (24), $\beta > 0$ indicates that returns to search are falling over time, meaning that search productivity conditional on income is rising ($\frac{\partial a}{\partial \tau} > 0$).

For price indices, I follow Aguiar and Hurst (2007) and Kaplan and Menzio (2015) in constructing average deviations in the prices that each household pays for each UPC relative to the average transaction price for that UPC. That is,

$$\text{PriceIndex}_{i,t} = \sum_{k \in \text{Transactions}(i,t)} \lambda_{i,t,k} (\log p_{i,t,k} - \log \overline{p_{i,\text{UPC}(k)}}),$$

where $\text{Transactions}(i, t)$ is the set of transactions made by household i in year t , $\lambda_{i,t,k}$ is the

Table D3: Coefficient on interaction between search intensity and year.

Measure of search: Dep. var.: $PriceIndex_{i,t}$	Log Trips per \$1K				Log Stores Visited per \$1K			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Search Intensity \times Year	-0.001* (0.001)	-0.001* (0.001)	-0.001** (0.001)	-0.001** (0.001)	-0.003** (0.001)	-0.003** (0.001)	-0.003** (0.001)	-0.003** (0.001)
Income Level FE	Yes	Yes	Yes		Yes	Yes	Yes	
Demographic Controls	Yes	Yes	Yes		Yes	Yes	Yes	
Year FE		Yes				Yes		
County-Year FE				Yes				Yes
N	917 634	917 634	917 634	917 634	917 634	917 634	917 634	917 634
R ²	0.07	0.08	0.08	0.15	0.08	0.10	0.10	0.17

Note: Table reports estimates of β from specification (25). Sample includes all NielsenIQ Homescan panelists from 2004–2019. Standard errors two-way clustered by household county and year. * indicates significance at 10%, ** at 5%.

expenditure share of transaction k in household i 's total expenditures in year t , $p_{i,t,k}$ is the price paid by household i , and $\overline{p_{t,\text{UPC}(k)}}$ is the average price paid for the UPC purchased in transaction k across all observed purchases in year t . In words, $PriceIndex_{i,t}$ captures whether household i pays a premium on average for the products it purchases. (I repeat the analysis constructing price indices using deviations in unit prices that households pay for a given brand or a given product category relative to all brand/product category purchases and find similar results.)

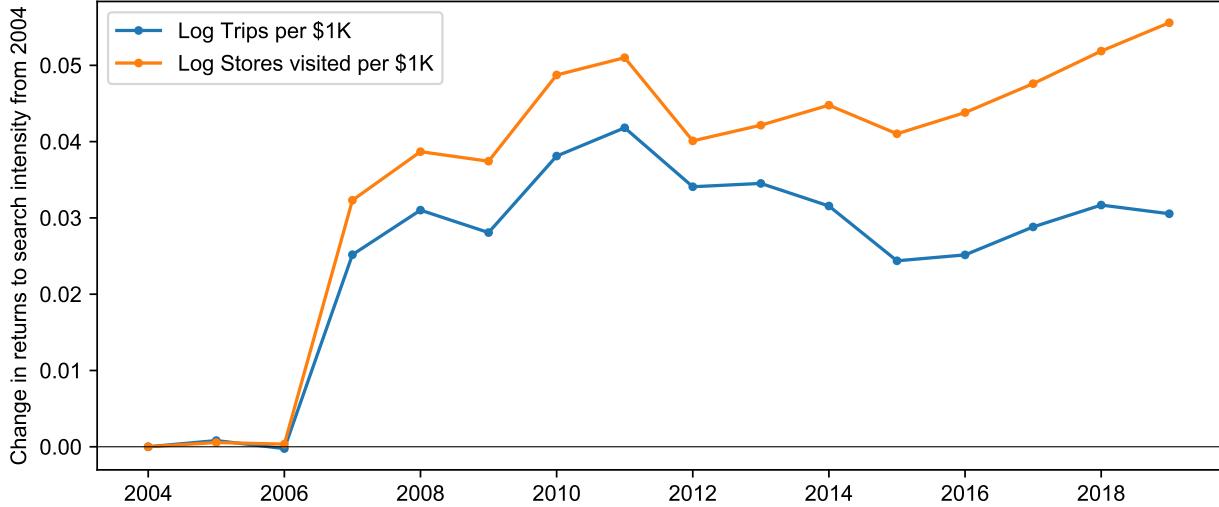
Table D3 reports the results, using number of trips and number of unique stores visited per \$1,000 expenditures as measures of search intensity. The estimated coefficient β is mildly negative, suggesting in fact that search productivity conditional on income is mildly decreasing over time ($\frac{\partial \alpha}{\partial t} < 0$). The results are robust to adding various fixed effects for household income levels, demographic characteristics, year, and county.

To visualize how returns to search evolve over time, I estimate a version of (25) with year dummies,

$$\begin{aligned} PriceIndex_{i,t} = & \alpha SearchIntensity_{i,t} + \sum_{t'} \beta_{t'} (SearchIntensity_{i,t} \times 1\{t = t'\}) \\ & + \gamma (SearchIntensity_{i,t} \times LogIncome_{i,t}) + \varphi_{\text{County}(i),t} + \psi' X_{i,t} + \varepsilon_{i,t}. \end{aligned}$$

This specification uses the most conservative set of controls from Table D3, including fixed effects for income levels and demographic characteristics (included in $X_{i,t}$), and county-year ($\varphi_{\text{County}(i),t}$). I omit a dummy for the year 2004, so that changes in returns to search

Figure D2: Returns to search, conditional on income, over time.



over time are measured relative to 2004.

Figure D2 plots $-\beta_t$ for this specification using two measures of search intensity. (I plot $-\beta_t$ since a negative coefficient β_t corresponds to an increase in the returns to search.) Both measures indicate returns to search rose sharply from 2006 to 2007, and then were flat or rose mildly thereafter.⁵⁴ These results again suggest that returns to search conditional on income are flat or slightly rising, indicating that search productivity conditional on income is flat or slightly falling over time.

D.2.2 Trends in shopping time

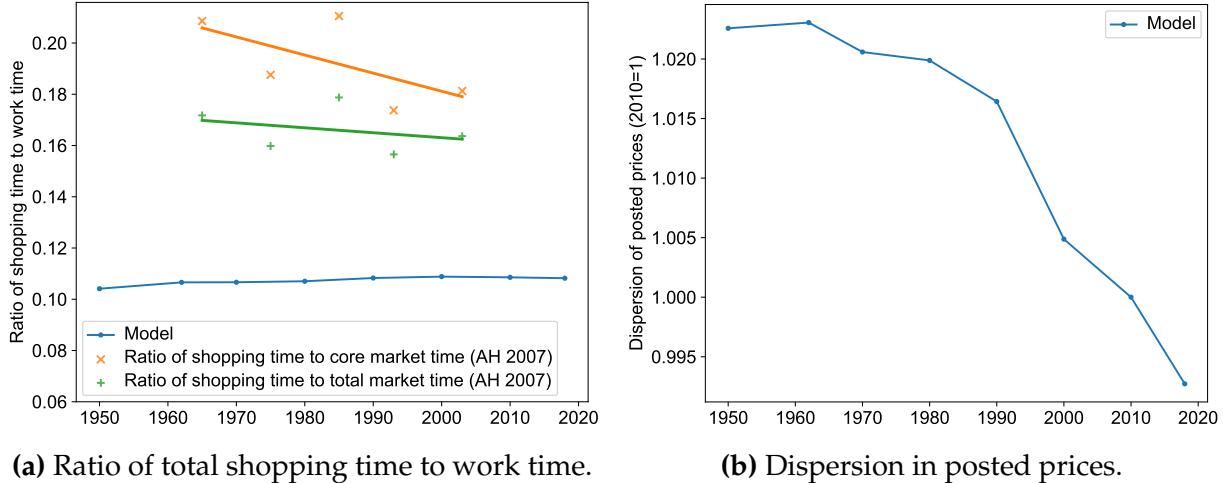
Döpper et al. (2021) show that both coupon redemption rates and time spent shopping for consumer products (in the American Time Use Survey) are declining over time. Döpper et al. (2021) interpret these measures as evidence of falling price sensitivity.

Interpreting trends in shopping time as evidence for changes in price sensitivity is challenging because improvements in search productivity may allow households to achieve the same search intensity while spending less time shopping. In some cases, an increase in search productivity can simultaneously result in a decrease in shopping time per purchase and an increase in search intensity.

To see this formally, using $s = at$ and taking the partial derivative of (23) with respect

⁵⁴To ensure that discontinuities from 2006 to 2007 do not bias the results in Table D3, I replicate Table D3 using data only from 2007–2019. While the statistical significance of some estimates are reduced, no specification produces a significant positive estimate for β .

Figure D3: Shopping time and price dispersion in the model.



to search productivity yields,

$$\frac{\partial t}{\partial a} = t \left(\frac{-p_s}{sp_{ss}} - 1 \right).$$

If the price function is sufficiently convex ($\frac{sp_{ss}}{-p_s} > 1$), then time per purchase can decrease with search productivity even as search intensity increases. Thus, it is difficult to conclude whether declines in shopping time per purchase indicate falling search intensity or rising search productivity.⁵⁵

D.2.3 Model predictions on search behavior and price dispersion

How do the evolution of shopping time and price dispersion in the model compare with the data? Figure D3a plots the ratio of total time spent shopping to total time spent working in the model from 1950 to 2018. This ratio is essentially flat over time in the model. For comparison, Figure D3a also plots these ratios using average weekly hours spent working and spent shopping for goods and services from Aguiar and Hurst (2007b). These data suggest the ratio of time spent shopping to time spent working is slightly declining or flat. Relative to the data, the level of shopping time is lower in the model and there is no declining trend. (As discussed above, while it is tempting to conclude that declining

⁵⁵A similar problem also affects the total amount of time allocated to shopping. In the model, the ratio of total time spent shopping to time spent working is equal to the elasticity of the price function, which is non-monotonic in income:

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

search time in the data suggests price sensitivity has fallen more than the model suggests, it is difficult to tease apart trends in search intensity from changes in search productivity with time series data on shopping time.)

Figure D3b plots average price dispersion in the model from 1950 to 2018. The model predicts a mild decline in the standard deviation of posted prices (less than 3 percent from 1950 to 2018). 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' search 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. 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.

Appendix E Comparison to Other Markup Measures

E.1 Production Function Estimation

De Loecker et al. (2020) show that accounting data can be used to estimate markups using what they call a “production function approach.” They apply this methodology to calculate markups for public firms in Compustat and find that average markups rose dramatically from 1980 to 2016.⁵⁶

It is challenging to map markups estimated using the production function approach to the retail markups used in this paper for two reasons: (1) these markups are calculated at the firm level, rather than at the product level, and (2) retailers’ identities are anonymized in the NielsenIQ data. However, in this section, I show markups estimated using the production function approach are also positively associated with buyer income.

Baker et al. (2020) use transaction-level data on debit and credit card spending by two million individuals to construct distributions of buyers’ incomes for a number of public firms from 2010–2015. Using their data, I calculate the average buyer income for each firm in each year by taking the expenditure-weighted average over the distribution of buyers’ incomes in their data. I use replication data from De Loecker et al. (2020) to calculate markups using the production function approach.⁵⁷ Merging the two datasets,

⁵⁶Bond et al. (2021) provide a useful guide of identification issues that make estimating markups using accounting data challenging. A salient concern is that revenue is often used as a proxy for output quantity (since output prices are not available in accounting data). Conflating revenue and output quantity poses important challenges for the identification of markups.

⁵⁷For comparison with average markups in the main text, the sales-weighted average markup for all

data on both production function markups and average buyer income are available for 378 firms, including 192 retail firms (NAICS industries 44–45). I also merge data on the sales shares of the top four, eight, twenty, and fifty firms in a NAICS-6 industry from the 2012 Economic Census, which is available for 152 of the 192 retail firms.

To measure the elasticity of the production function markup to buyer income, I estimate the specification,

$$\log(\text{Markup})_{fit} = \alpha_{it} + \beta \log(\text{Avg. Buyer Income})_{fit} + \varepsilon_{fit},$$

where f indexes firms, i indexes NAICS-4 industries, t indexes years, and α_{it} are year-NAICS-4 fixed effects.

Table E1 reports the results. In column 1, the estimated elasticity using all firms in the merged dataset is 23 percent. Since the results in the main text are concerned with retail markups, in columns 2–6, I limit my analysis to firms in NAICS industries 44–45. For this subsample, the estimated elasticity of production function markups to buyer income is 36 percent (column 2).

Columns 3–6 add controls for sales concentration in NAICS-6 industries, using data from the 2012 Economic Census. Once these controls are included, the estimated elasticity of production function markups to buyer income is slightly higher (44 percent). However, these measures of sales concentration are not significantly associated with retail firm markups.

E.1.1 Upstream firm markups and downstream buyer income

The analysis in the main text opens the question of whether changes in consumer behavior could be responsible for changes in markups in other sectors besides retail. Theory suggests that changes in price sensitivity downstream could affect markups upstream. For example, in standard models of vertical supply chains (e.g., Tirole 1988 Ch. 3), a reduction in the elasticity of consumer demand results in a uniform increase in markups along the entire chain of producers. Wu (2022) shows that similar intuitions may hold in a general production network.

I use the De Loecker et al. (2020) markups to provide suggest evidence for this channel in the data. In particular, I show that the markups of upstream firms are increasing in the average income of buyers at downstream firms they supply to.

I begin by constructing a sample of matched upstream–downstream firm pairs using

firms in 2007 in the De Loecker et al. (2020) replication code is 1.47, and the cost-weighted average for all firms in 2007 is 1.27.

Table E1: Impact of buyer income and sales concentration on De Loecker et al. (2020) markups.

<i>Log Production Function Markup</i>	All	Retail Firms (44–45)				
	(1)	(2)	(3)	(4)	(5)	(6)
Log Avg. Buyer Income	0.229** (0.085)	0.358** (0.067)	0.439** (0.094)	0.444** (0.094)	0.443** (0.094)	0.444** (0.095)
Top 4 Firms Sales Share			-0.101 (0.123)			
Top 8 Firms Sales Share				-0.067 (0.117)		
Top 20 Firms Sales Share					-0.082 (0.136)	
Top 50 Firms Sales Share						-0.071 (0.139)
Year × NAICS-4 FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	1706	898	693	693	693	693
R ²	0.76	0.71	0.68	0.68	0.68	0.68
Within R ²	0.02	0.17	0.21	0.21	0.21	0.21

Note: Firm production function markups are calculated using replication code from De Loecker et al. (2020). Average buyer income is calculated using data on the income distribution of each firm's customers from Baker et al. (2020). The sales shares of the top 4, 8, 20, and 50 firms in a NAICS-6 industry are from the 2012 Economic Census. All regressions weighted by consumer spending from Baker et al. (2020), and standard errors are two-way clustered by firm and year. * indicates significance at 10%, ** at 5%.

Table E2: Relationship between buyer income at downstream firms with markups of upstream suppliers.

<i>Markup at Upstream Firm</i>	(1)	(2)	(3)	(4)
Log Avg. Buyer Income of Downstream Firm	0.103 (0.119)	0.078** (0.033)	0.085** (0.031)	0.076** (0.033)
Year FE _s	Yes	Yes	Yes	Yes
Year-Upstream Industry FE _s		Yes	Yes	
Year-Downstream Industry FE _s			Yes	
Year-Upstream Industry-Downstream Industry FE _s				Yes
<i>N</i>	9092	8919	8484	7765
<i>R</i> ²	0.00	0.74	0.76	0.80

Note: Firm production function markups are calculated using replication code from De Loecker et al. (2020). Average buyer income is calculated using data on the income distribution of each firm's customers from Baker et al. (2020). Upstream and downstream firms are matched using data from Compustat Customer Segments. * indicates significance at 10%, ** at 5%.

data from Compustat Customer Segments. These data are compiled from firm disclosures about geographies, industries, or customers that comprise over 10 percent of their sales, and have been previously used by Cohen and Frazzini (2008) to identify links between firms. I restrict the sample to cases where the downstream customer is a single firm and construct a manual crosswalk from downstream firm names in the Compustat Customer Segments data to firm names used by Baker et al. (2020). Finally, I merge these linked firm pairs with data on markups of public firms estimated by De Loecker et al. (2020). The final sample includes public firms from 2010 to 2015, with markups on the upstream firm in each year available from 2010 to 2015 and the average buyer income at the downstream firm from each year available from 2010 to 2015.

Table E2 reports how markups at the upstream firms relate to average buyer income at downstream firms. The unconditional relationship is positive but insignificant. Controlling for the industry of the upstream firm, the industry of the downstream firm, or both leads to a significant positive elasticity of upstream markups to income at downstream firms. To take an example, within a given upstream industry, doubling the income of buyers at an important retail customer is associated with an 8 percent increase in the markup of the upstream firm.

These regressions measure correlations in the cross-section of firms, and of course do not isolate causal effects. However, they are consistent with models where changes in income or price sensitivity downstream can affect the markups charged by upstream

firms.

E.2 Demand Estimation

A vast literature in industrial organization estimates structural models of demand to recover marginal costs and markups from data on prices. In this section, I estimate elasticities of demand and markups in a single product module (margarine). I then compare the marginal costs and markups recovered from demand estimation to the retail markups that I use in the body of the paper. I find that unit marginal costs and markups recovered from demand estimation are positively correlated with unit wholesale costs and retail markups. The markups recovered from demand estimation are larger in magnitude and move more than one-for-one with retail markups.

When estimating markups in this setting, it is important to specify the form of conduct between manufacturers and retailers, since different forms of conduct affect whether the markups recovered from demand estimation are comparable to the retail markups used in the main text.⁵⁸ For this exercise, I assume that manufacturers and retailers maximize joint surplus and then divvy that surplus according to (unmodeled) two-part tariffs or rebate arrangements. This assumption on conduct means that retail margins should form part of the overall markup recovered from demand estimation.

I start by describing how I construct data on margarine prices and market shares, as well as cost shifters and market sizes.⁵⁹ I then specify and estimate a random coefficients discrete choice model of demand à la Berry et al. (1995). Finally, I compare how marginal costs and markups recovered from demand estimation compare to the wholesale costs and retail markups I use in the main text.

Data construction. For data on prices and market shares, I use NielsenIQ Retail Scanner data from 2006 to 2009, which includes average weekly prices and total sales of margarine at each store participating in NielsenIQ Retailer Scanner program. I define geographic markets using designated market areas (DMAs) provided by NielsenIQ. These are non-

⁵⁸For example, Butters et al. (2022) suggest that retailers use simple cost-plus pricing rules. In this case, retailer margins would appear as part of manufacturers' marginal costs recovered from demand estimation, rather than as part of markups, as noted by Döpper et al. (2021).

⁵⁹I choose margarine as the product module for demand estimation since the PromoData has especially high coverage in margarine: more than 70 percent of margarine sales in the Homescan panel in 2007 are matched to wholesale costs in PromoData. One concern is that margarine may be somewhat durable (opened margarine has a shelf life of about 2–3 months), though previous work has considered consumer food products with longer shelf lives (e.g., Nakamura and Zerom 2010 estimate a similar demand system in the market for ground coffee, which has a shelf of 3–5 months). For this reason, I aggregate prices and shares at the monthly level, rather than using the weekly data provided by NielsenIQ.

overlapping groups of counties that are comparable in size to CSAs. For parsimony, I consider the 40 DMAs with the largest volume of margarine sold from 2006 to 2009: these markets account for over half of the total volume of margarine sold in the NielsenIQ Retail Scanner data over this period. I aggregate sales and average unit prices of each margarine UPC sold by each retail chain in each DMA at the monthly level. This yields $40 \times 12 \times 4 = 1920$ distinct markets for our estimation exercise.

For consumer demographics in each market, I use data on NielsenIQ Homescan panelists from 2006 to 2009. Since the NielsenIQ Homescan data records the DMA in which households reside, these demographic measures are consistent with the DMA boundaries that I use to define markets.⁶⁰ I collect the demographic information for panelists in each year in each DMA and use the projection weights provided by NielsenIQ to weight panelists.

I define a product as a unique retailer-UPC combination. As noted by Broda et al. (2009), retail chains offer different shopping amenities, and hence consumers may see the same UPC offered at two different retail chains as distinct products. One concern is that retailers and products with small sales and quantities purchased may have lower-quality data. Hence, I limit my analysis to the largest forty retailers by sales in the sample. I also isolate the top 150 UP Cs by total sales and combine the remaining products with the outside good. This approach allows me to reduce the number of products considered for demand estimation without aggregating distinct UP Cs and losing product-specific attributes. The top 150 UP Cs account for over 80 percent of the total volume of margarine sold in the DMAs in my sample. The final sample includes 338,726 observations for prices and volume across 40 DMAs in 48 months.

I construct ownership matrices using brand identifiers from NielsenIQ. The UP Cs in my sample belong to 20 distinct brands. An alternative is to use data on company ownership from Global Standards One (GS1), which identifies the parent company that owns each UPC (see Appendix A). Using GS1 parent companies to construct ownership matrices does not qualitatively change the results I present here.

Since prices are endogenous, estimating the demand system requires instruments that are orthogonal to movements in demand. Following Villas-Boas (2007), I use input prices as instruments for prices. In particular, I use monthly data on the price of soybeans, corn, oil-producing crops, and milk from the USDA National Agricultural Statistics Service (NASS) from 2006–2009 as cost shifters. Margarine is mostly made from vegetable oils, including corn and soy; and butter, a primary alternative to margarine, is made from

⁶⁰An alternative is to use demographic data from the American Community Survey (ACS) or another external source and build a crosswalk from geographies in the ACS to NielsenIQ's DMAs.

cream or milk. The exclusion restriction is that these agricultural commodity costs are not influenced by trends in demand for margarine, which is plausible given that a vast number of other food products require soybeans, corn, and milk as inputs.

To define the size of the market and the market share of the outside option, I use the NielsenIQ Homescan data to calculate average annual consumption of margarine per household in each DMA. The median across all DMAs and time periods is about 9 pounds of margarine per household per year. Hence, I take the market size in each DMA in each month to be 0.75 pounds of margarine per household, times the number of households in the DMA. Constructed in this manner, the market share of the outside option displays considerable variation across time and space: it ranges from just over 10 percent to over 95 percent across the markets and months considered.

Demand model. I estimate a random coefficients discrete choice model of demand (Berry et al. 1995). I assume that the utility received by a household i for purchasing a (retailer-UPC) product j in DMA m at time t is

$$u_{ijmt} = \alpha_i^0 + \alpha_i^p p_{jmt} + \delta_{jmt} + \varepsilon_{ijmt}, \quad (26)$$

where α_i^0 is household i 's mean taste for purchasing any product, α_i^p determines household i 's price sensitivity, p_{jmt} is the price of product j in market m at time t , δ_{jmt} is an unobserved demand shifter that varies across products, markets, and time periods, and ε_{ijmt} is an idiosyncratic draw from a Gumbel distribution. (When estimating the demand model, I include UPC-retailer fixed effects, which absorb product-specific characteristics.) Households in market m and time t purchase exactly one unit of the product that gives them the highest utility (which may be the outside option).

I assume that the coefficients α_i^0 and α_i^p are given by

$$\begin{aligned} \alpha_i^0 &= \bar{\alpha}^0 + \hat{\alpha}_i^0 + \Pi_{z0} z_i, \\ \alpha_i^p &= \bar{\alpha}^p + \hat{\alpha}_i^p + \Pi_{zp} z_i, \end{aligned}$$

where $\bar{\alpha}^0$ and $\bar{\alpha}^p$ are means across households, z_i is log of household i 's income, Π_{z0} and Π_{zp} are interactions that allow the coefficients to vary systematically with income, and where $\hat{\alpha}_i^0$ and $\hat{\alpha}_i^p$ are random draws from normal distributions with variance Σ_0 and Σ_p ,

$$\hat{\alpha}_i^0 \sim \mathcal{N}(0, \Sigma_0), \quad \hat{\alpha}_i^p \sim \mathcal{N}(0, \Sigma_p).$$

These random coefficients allow more flexibility for the model to match patterns of

Table E3: Demand system estimates.

	(1) Logit	(2) Random coefficients
Price	-1.063 (0.014)	-5.038 (0.150)
Π_{z0}		-0.375 (0.867)
Π_{zp}		3.259 (0.133)
Median price elasticity	1.91	2.30
Median estimated markup	1.81	1.78
UPC \times Retailer FE	Yes	Yes
DMA FE	Yes	Yes
N	338726	338726

Note: Reported coefficients are estimated using the `pyblp` package developed by Conlon and Gortmaker (2020), with bootstrapped standard errors in parentheses. Column 1 reports results from the logit model specification in (27), and column 2 reports results from the random coefficients specification described in (26).

substitution across products. Moreover, the parameters Π_{z0} and Π_{zp} allow the utility from purchasing margarine and price sensitivity to vary with income. Admittedly, one could choose a richer set of interactions—for example, by allowing the price coefficient to also be interacted with age and household size or by allowing heterogeneity in demand for other product characteristics. For parsimony, I focus on only the income interactions.

Results. I use the `pyblp` package developed by Conlon and Gortmaker (2020) to estimate the demand system. To gain intuition, before estimating the random coefficients model, I start by estimating a logit model with the specification,

$$\log s_{jmt} = \alpha^0 - \alpha^p p_{jmt} + \gamma_j + \phi_m + \delta_{jmt}, \quad (27)$$

where s_{jmt} is the market share of product j in market m at time t , γ_j are product (retailer-UPC) fixed effects, which absorb other product characteristics, and ϕ_m are DMA fixed effects.

Column 1 of Table E3 reports the results from estimating this logit model. The estimated coefficient on price α^p is negative, yielding a median price elasticity of demand of 1.91. As a result, the median markup across all products is 1.81.

Column 2 of Table E3 reports results from estimating the random coefficients model. The estimated coefficient on price remains significantly negative. The interaction of the value of the outside option with income, Π_{z0} , is slightly negative but not significant. However, the estimated interaction of the price coefficient with income, Π_{zp} , is significantly positive, suggesting that high-income consumers are less price sensitive.⁶¹ The median price elasticity of demand from the random coefficients model is moderately higher than in the logit model, at 2.30. This estimate appears to match well with previous studies of the margarine market: Kim (2008) estimates a demand system in the margarine market and finds that price elasticities range from 1.85 to 6.34, and Griffith et al. (2010) estimate a demand system in the butter and margarine market and find an average price elasticity of 2.44.

I now turn to measuring how marginal costs and markups recovered from demand estimation compare to the wholesale costs and retail markups that I use in the main text. For each UPC in each month, I calculate the median marginal cost estimated using the random coefficients model. In 8 percent of cases, the median marginal cost is negative, and in 17 percent of cases the median marginal cost implies a markup of over 5. I trim these observations and merge the remaining monthly marginal cost estimates into the consumer purchase data, yielding 194,582 consumer transactions that are matched to estimates of marginal cost and 155,856 consumer transactions where both the PromoData wholesale cost and the demand estimation-based measure of marginal cost are available.

Table E4 shows how unit marginal costs and markups compare for the sample of margarine purchases where both the wholesale cost data and demand estimates are available. Wholesale costs comove closely with marginal costs recovered from demand estimation. (Both are measured in dollars per pound of margarine, so that differences in package size across products do not affect the results). Accordingly, the retail markups and the markups recovered from demand estimation also exhibit a strong positive correlation ($\rho \approx 0.6$).

I explore the differences between unit marginal costs and markups recovered from demand estimation and the wholesale cost data in Tables E5 and E6. Column 1 of Table E5 shows that, on average, marginal costs of margarine recovered from demand estimation are about 40 cents per pound lower than the base wholesale costs in the PromoData and move slightly more than one-for-one with wholesale costs. Intuitively, marginal costs faced by margarine producers should be lower than the wholesale costs faced by retailers. Column 3 shows that marginal costs of margarine are also about 10 cents per pound lower

⁶¹The estimation procedure also yields estimates of Σ_0 and Σ_p , which allow for idiosyncratic variation in the utility of purchasing margarine and in price sensitivity across households. The estimates of Σ_0 and Σ_p are not significantly different from zero.

Table E4: Correlation coefficients for unit marginal costs and markups measured from demand estimation versus wholesale cost data.

Correlation coefficient (ρ)	<i>Demand estimation</i>	
	Log(Unit Marginal Cost)	Log(Markup)
Log(Unit Wholesale Cost)		
Using PromoData base price	0.88	
Using PromoData deal price	0.84	
Log(Retail Markup)		
Using PromoData base price		0.58
Using PromoData deal price		0.63

Table E5: Relationship between estimated marginal costs and wholesale costs.

<i>Unit marginal cost (demand estimation)</i>	(1)	(2)	(3)	(4)
PromoData unit wholesale cost (base)	1.087** (0.032)	0.432** (0.133)		
PromoData unit wholesale cost (deal)			0.998** (0.050)	0.345** (0.101)
Constant	-0.395** (0.058)		-0.098 (0.067)	
UPC FEs		Yes		Yes
N	155856	155856	155856	155856
R ²	0.92	0.98	0.87	0.98

Note: Marginal costs and wholesale costs are expressed in dollars per pound of margarine, so that differences in package sizes across products do not affect results. Standard errors are clustered by brand. * indicates significance at 10 percent, ** at 5 percent.

than the deal wholesale costs in the PromoData, though the constant is not significant. In columns 2 and 4, I also show that within-UPC variation in wholesale costs by month also predicts within-UPC variation in estimated marginal costs by month. Accordingly, Table E6 shows that the markups recovered from demand estimation are 35–50pp higher than retail markups on average and move slightly more than one-for-one with differences in retail markups across products and over time.

Table E6: Relationship between estimated markups and retail markups.

<i>Markup (demand estimation)</i>	(1)	(2)	(3)	(4)
Retail markup (using base wholesale cost)	1.147** (0.097)	1.356** (0.078)		
Retail markup (using deal wholesale cost)			0.874** (0.111)	1.099** (0.062)
Constant	0.358** (0.181)		0.459** (0.159)	
UPC FEs		Yes		Yes
N	155856	155856	155856	155856
R ²	0.34	0.89	0.40	0.89

Note: Standard errors are clustered by brand. * indicates significance at 10 percent, ** at 5 percent.

Appendix F More Facts about Retail Markups

In this appendix, I document stylized facts from the data that are tangential to the paper’s core argument but may be pertinent to other work. Section F.1 documents how retail markups covary with income and with characteristics such as market share, retailer size, and product quality. Section F.2 explores the “morphology” of markup dispersion, borrowing techniques from Kaplan and Menzio (2015). Finally, Section F.3 explores the degree of segmentation in buyers’ incomes across retail chains.

F.1 Markups across products

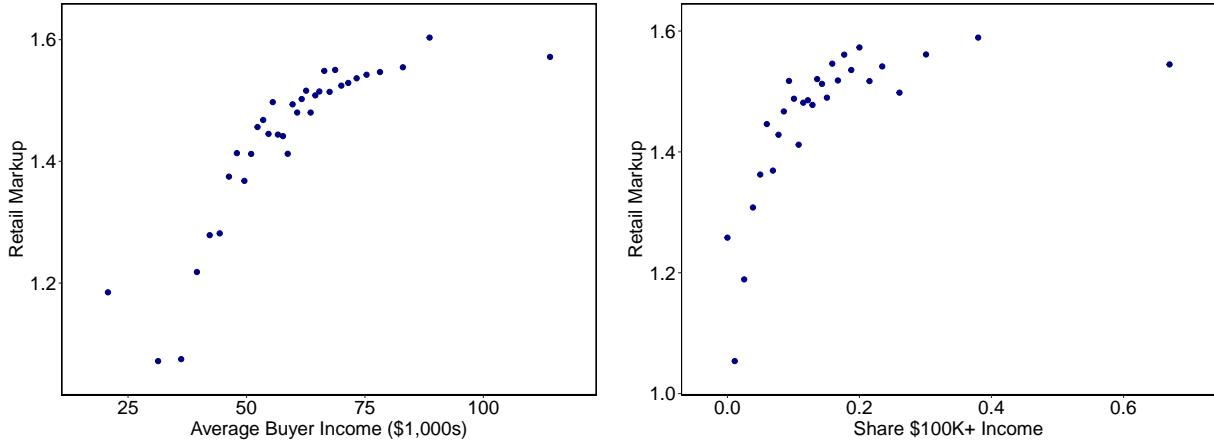
This section catalogues five stylized facts about how retail markups in the sample covary with product and firm characteristics.

1. Products with a high-income customer base have higher retail markups, unconditionally, within manufacturing firm, and within retailer. Figure F1 shows a binscatter of UPC retail markups against two measures of buyer income—the sales-weighted average income of buyers across all observed transactions, and the share of purchases from households with over \$100K in income. In magnitudes, doubling average buyer income is associated with a 6.5 percent increase in retail markups, increasing the share of buyers with over \$100K in income 10pp is associated with a 9.2 percent increase in retail markups, and increasing the share of buyers with under \$50K in income 10pp is associated with a 5.1 percent decrease in retail markups.

The covariance of retail markups and buyer income remains positive for products that are produced by the same manufacturing firm. I identify manufacturing firms using a crosswalk from UPCs to parent companies provided by GS1 (see Appendix A). Figure F2 shows that retail markups increase with buyer income for two products manufactured by the same parent company, unconditionally and within product module. The within-store results in the main text (Figure 2b) show that retail markups also covary positively with buyer income within retail chains. Note that these results contrast with predictions of standard representative agent macro models, such as the one developed in Hottman et al. (2016), that predict that markups vary only at the firm level.

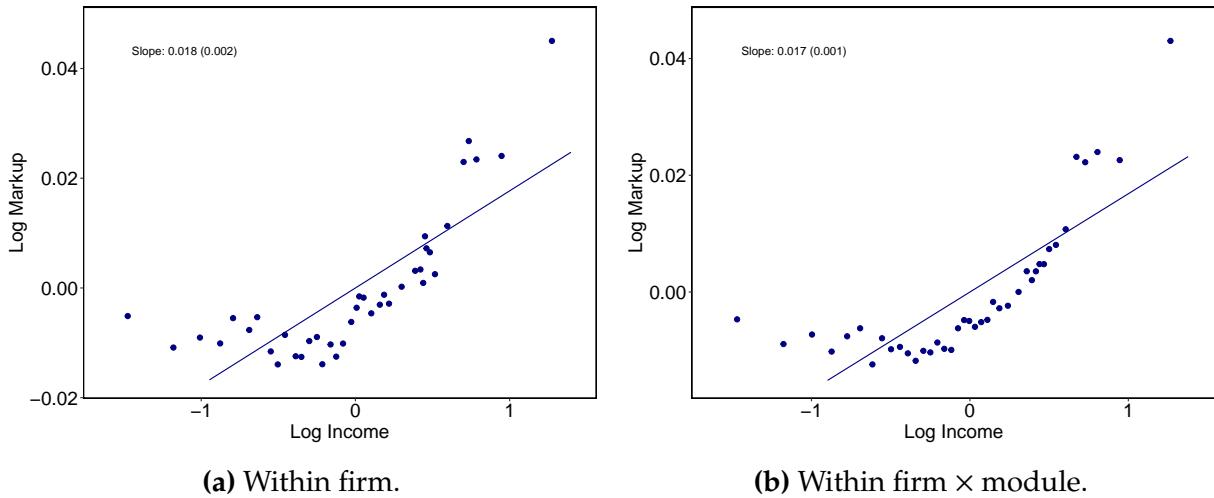
2. Larger retailers charge lower markups on average. Figure F3 plots the relationship between retail markups and retailer size (measured by total expenditures at the retailer in the NielsenIQ Homescan data), unconditionally and for identical products. Doubling

Figure F1: UPC retail markup and buyer income.



Note: UPC retail markups are calculated as sales-weighted averages over all observed transactions. Graphs show a binned scatter weighted by UPC sales.

Figure F2: Retail markups increase with income within manufacturing firm.



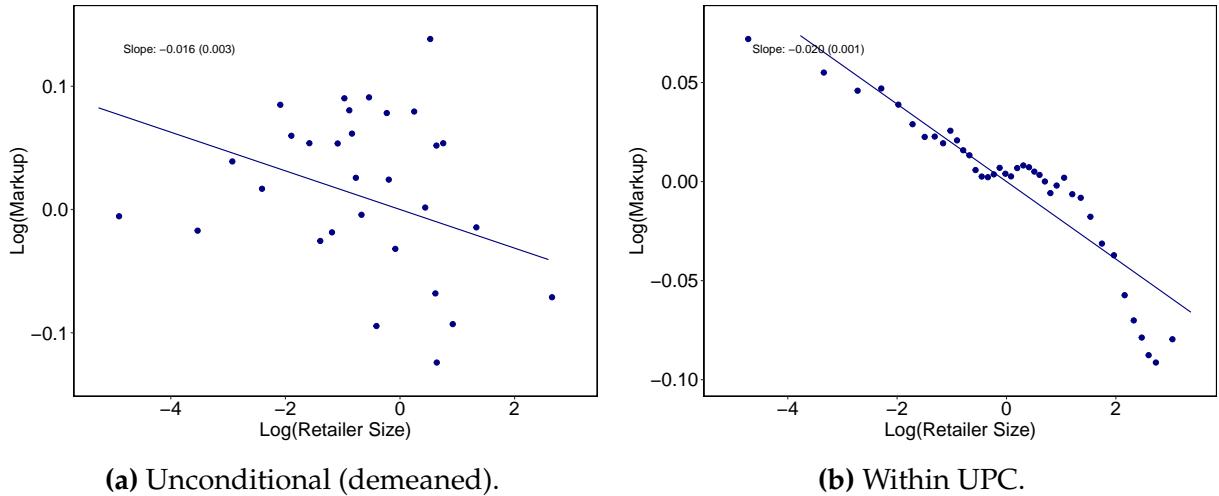
Note: Graphs show a binned scatter weighted by sales.

retailer size is associated with 1.6 percent lower markups overall and with 2.0 percent lower markups on identical products.

When comparing markups across retailers of different sizes, the concerns about mismeasurement of marginal costs due to unobserved volume discounts or retailer rebates discussed in Section 3.1.1 are especially salient. I undertake two robustness checks. First, I compare how the elasticity of retail markups to size varies as we remove the largest retailers from the sample. If volume discounts are concentrated among the largest retailers,

and the negative relationship between markups and retailer size is indeed driven by mis-measurement, the relationship between retailer size and markups should disappear as we remove the largest retailers from the sample. Table F1 shows that retailer size continues to be associated with lower markups as we remove large retailers from the sample (though the elasticity attenuates slightly from 2.0 percent to 0.8–1.1 percent). Second, I check the elasticity of De Loecker et al. (2020) markups of public retail firms to retailer size. Table F2 shows that while markups positively correlate with firm size in the sample of all public firms, markups indeed fall with size for retail firms. The elasticity of markups to firm size is in line with the elasticity in my microdata.

Figure F3: Larger retailers charge lower markups.



Note: Graphs show a binned scatter weighted by sales.

Table F1: Robustness of relationship between retail size and markups.

	Excluding top retailers by size						
<i>Log Retail Markup</i>	All (1)	Top 1 (2)	Top 3 (3)	Top 5 (4)	Top 10 (5)	Top 50 (6)	Top 100 (7)
Log Retailer Size	-0.020** (0.001)	-0.011** (0.001)	-0.011** (0.001)	-0.009** (0.001)	-0.008** (0.001)	-0.021** (0.002)	-0.008** (0.002)
UPC FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N (millions)	25.0	19.9	18.3	17.5	14.0	5.4	2.4
R ²	0.61	0.59	0.59	0.58	0.59	0.63	0.70

Note: Regression weighted by sales, and standard errors two-way clustered by product brand and household county. * indicates significance at 10%, ** at 5%.

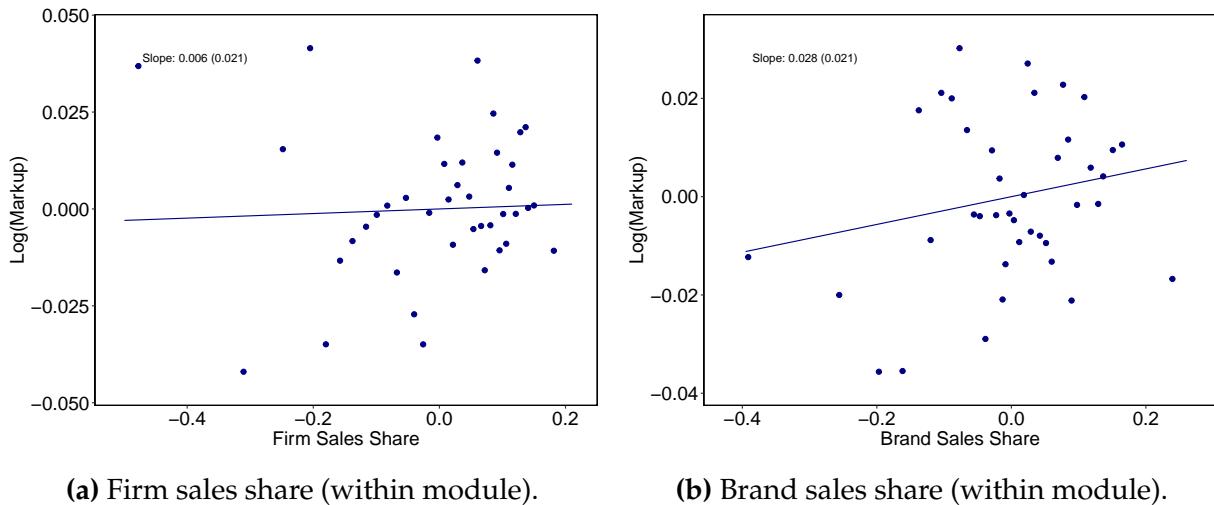
Table F2: Larger retail public firms have lower De Loecker et al. (2020) markups.

<i>Log Production Function Markup</i>	All firms		Retail firms (NAICS 44–45)			
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	0.002 (0.006)	0.026** (0.007)	-0.026** (0.005)	-0.025** (0.005)	-0.016** (0.005)	-0.017** (0.006)
Year FE _s	Yes	Yes	Yes	Yes	Yes	Yes
Year \times 2-digit NAICS FE _s		Yes		Yes	Yes	Yes
Year \times 3-digit NAICS FE _s		Yes			Yes	Yes
Year \times 4-digit NAICS FE _s		Yes				Yes
N	209479	207749	11960	11960	11912	11767
R ²	0.00	0.24	0.06	0.07	0.17	0.22

Note: Production function markups are calculated using replication code from De Loecker et al. (2020). Sample includes retail firms (NAICS 44 or 45) from 1980 to 2016. Standard errors two-way clustered by year and 4-digit NAICS industry. * indicates significance at 10%, ** at 5%.

3. There is no strong relationship between firm/brand sales share and retail markups. Figure F4 plots the relationship between the sales share of a parent company (identified by GS1) and of a brand (using brand identifiers from NielsenIQ) with retail markups, controlling for product module. The relationship between both measures of sales share and retail markups is flat or slightly positive but statistically insignificant.

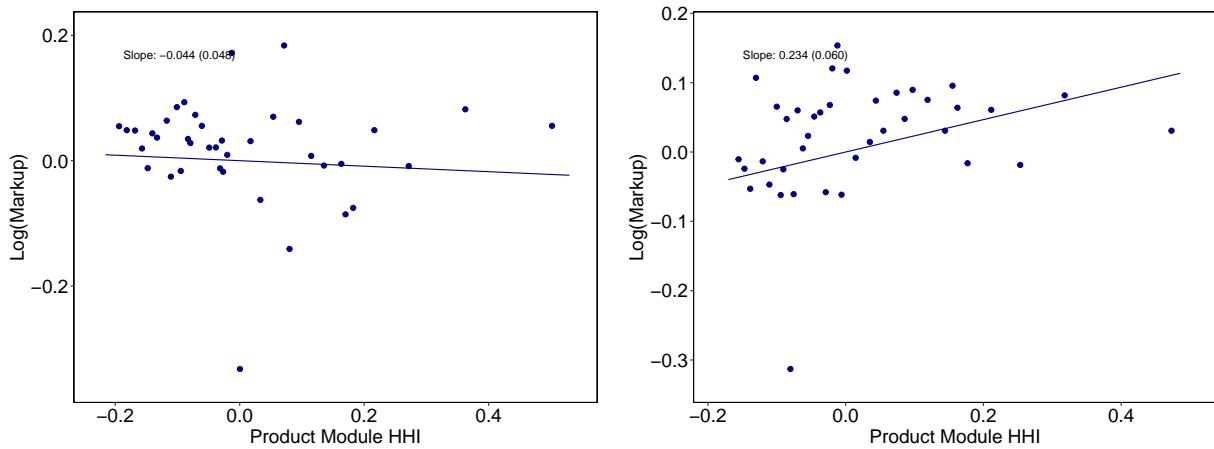
Figure F4: Relationship between retail markups and sales share.



Note: Graphs show a binned scatter weighted by sales.

4. There is no strong relationship between sales concentration and retail markups. Figure F5 plots the relationship between retail markups and the Herfindahl-Hirschman index (HHI) of concentration in the product module. The left panel measures sales shares at the parent company level, while the right panel measures sales shares at the brand level. The resulting relationship between retail markups and HHIs is flat or only mildly positive. Appendix Table E1 also shows that the relationship between De Loecker et al. (2020) markups of public retail firms and sales concentration is also insignificant after accounting for buyer income. These findings echo the point made by Berry et al. (2019) that reduced form regressions of markups on concentration are not necessarily informative about causal effects.

Figure F5: Relationship between retail markups and HHI of product module.



(a) HHI using firm sales shares (demeaned). **(b)** HHI using brand sales shares (demeaned).

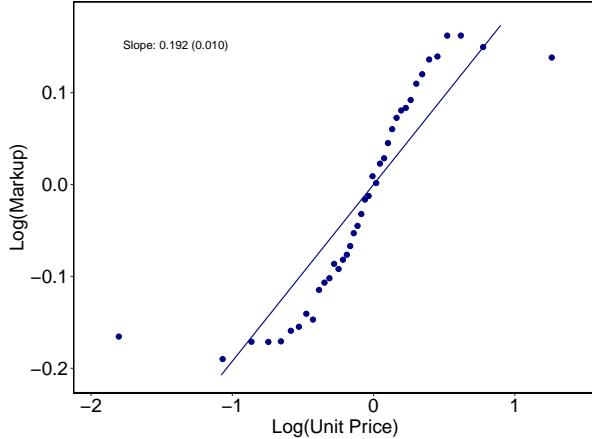
Note: Graphs show a binned scatter weighted by sales.

5. More expensive products within a product module have higher retail markups. Figure F6 plots the relationship between unit prices and retail markups. Within a product module, products with higher unit prices tend to have higher retail markups. If we take unit prices as a proxy for quality, this suggests that markups tend to increase with quality.

F.2 The morphology of markup dispersion

This appendix section is named after an influential paper by Kaplan and Menzio (2015) that documents the sources of variation in prices of identical products. I follow their strategy to decompose the dispersion in markups in my sample.

Figure F6: Relatively expensive products have higher retail markups.



(a) Log unit price (within module).

Note: Graphs show a binned scatter weighted by sales.

F.2.1 Decomposing sources of markup variation across transactions

I utilize two different decompositions of the variation in markups observed across all transactions. The first uses the fact that the markup on good g sold at store s in transaction t can be decomposed as

$$\mu_{g,s,t} = \underbrace{\bar{\mu}}_{\text{Average markup}} + \underbrace{(\bar{\mu}_s - \bar{\mu})}_{\text{Store specific component}} + \underbrace{(\bar{\mu}_{g,s} - \bar{\mu}_s)}_{\text{Store-good specific component}} + \underbrace{(\mu_{g,s,t} - \bar{\mu}_{g,s})}_{\text{Transaction component}}.$$

That is, the markup for the transaction depends on how a store's average markup compares to the average markup across all stores, how the store's average markup for good g compares to the average markup it charges across all products, and how the markup on transaction t compares to the average markup paid for that good at the store. This decomposition is analogous to the one reported in Kaplan and Menzio (2015) Table 3.

Table F3 shows that this decomposition attributes about 5 percent of the variation in markups across transactions to the variation in the average markups charged by a store, 90 percent to differences in average markups charged for the different products at the same store, and 9 percent to differences in markups across transactions of the same good at the same store. If some store–good combinations are observed infrequently, this decomposition may attribute too much of the variation to the store–good specific component compared to the transaction component. Hence, I repeat this decomposition by instead aggregating average markups at the retail chain level. The amount of variation

Table F3: Decomposition of the variance of markups across transactions.

	Store-level	Retailer-level
Store/retailer specific component	5%	5%
Store/retailer-good specific component	87%	71%
Transaction component	9%	24%
Covariances	-3%	0%
Good specific component	61%	62%
Store/retailer-good specific component	29%	14%
Transaction component	9%	24%
Covariances	0%	0%

due to differences in markups across retail chains is about the same (consistent with the fact that retail chains tend to set uniform prices, as documented by DellaVigna and Gentzkow 2019), but the decomposition now attributes more of the variation in markups—about 25 percent—to idiosyncratic differences in markups paid across transactions at the same retail chain for the same good. These results are consistent with the findings in Kaplan and Menzio (2015) that differences in the average “expensiveness” of stores account for a relatively small amount of variation in prices or markups in the data.

The second decomposition instead decomposes the markup on good g sold at store s in transaction t as

$$\mu_{g,s,t} = \underbrace{\bar{\mu}}_{\text{Average markup}} + \underbrace{(\bar{\mu}_g - \bar{\mu})}_{\text{Good specific component}} + \underbrace{(\bar{\mu}_{g,s} - \bar{\mu}_g)}_{\text{Store-good specific component}} + \underbrace{(\mu_{g,s,t} - \bar{\mu}_{g,s})}_{\text{Transaction component}}.$$

Table F3 shows that this decomposition attributes about 60 percent of the variation in markups across transactions to the variation in average markups for a UPC, 30 percent to differences in average markups charged for the same product across stores, and 10 percent to differences in markups across transactions of the same good at the same store. Aggregating at the retailer level yields broadly similar results.

These results suggest that the majority of variation in markups comes from differences in average markups across products. This is consistent with the evidence presented in the main text that even conditional on shopping at a specific store, households from different income groups can pay different markups because of differences in the set of products that they purchase.

F.2.2 Decomposing sources of markup variation across households

Next, I consider how sources of variation in the aggregate markup paid by each household. The standard deviation of markups across households is 14.4pp. (Note that the markup gap across income groups in Section 3.1 is equal to just over one standard deviation in the distribution of markups paid across households.)

Following Kaplan and Menzio (2015), I decompose the aggregate markup paid by household i as

$$\bar{\mu}_i = \underbrace{\bar{\mu}}_{\text{Average markup}} + \underbrace{(\hat{\mu}_g - \bar{\mu})}_{\text{Good specific component}} + \underbrace{(\hat{\mu}_{g,r} - \hat{\mu}_g)}_{\text{Retailer-good specific component}} + \underbrace{(\bar{\mu}_i - \hat{\mu}_{g,r})}_{\text{Transaction component}},$$

where $\hat{\mu}_g$ is the aggregate markup household i would pay if they paid the average markup charged on all goods g in their consumption basket, and $\hat{\mu}_{g,r}$ is the aggregate markup household i would pay if they paid the average markup charged by the retailers they shop at for the goods they buy. In other words, the good specific component captures whether household i buys goods with high average markups, the retailer-good specific component captures whether household i buys those goods from retailers that have higher average markups, and the transaction component captures whether household i buys those goods at times when markups are relatively high.

Table F4 suggests all three components play a sizeable role in the variation in markups paid across households. Both the good specific and retailer-good specific components constitute about a third of the variation in markups across households, and the transaction component constitutes 15 percent of the variation. Moreover, all three components have a positive covariance: households that buy relatively high-markup products tend also to buy those products from high-markup retailers and buy at relatively high prices. These results are consistent with the way search behavior and tastes for quality vary systematically across households in the calibrated model in the main text. A decomposition that instead isolates the retailer specific component of variation in markups paid across households delivers similar results, as shown in Table F4.

F.3 Selection by income into retail chain

DellaVigna and Gentzkow (2019) find that prices and assortments are surprisingly uniform across stores within retail chains. They argue that uniform pricing can result in large losses relative to optimal prices because, within a retail chain, stores in high-income areas have higher-income consumers and thus are likely to face more inelastic demand curves.

This section considers the possibility that, due to self-selection of consumers into

Table F4: Decomposition of the variance of markups paid across households.

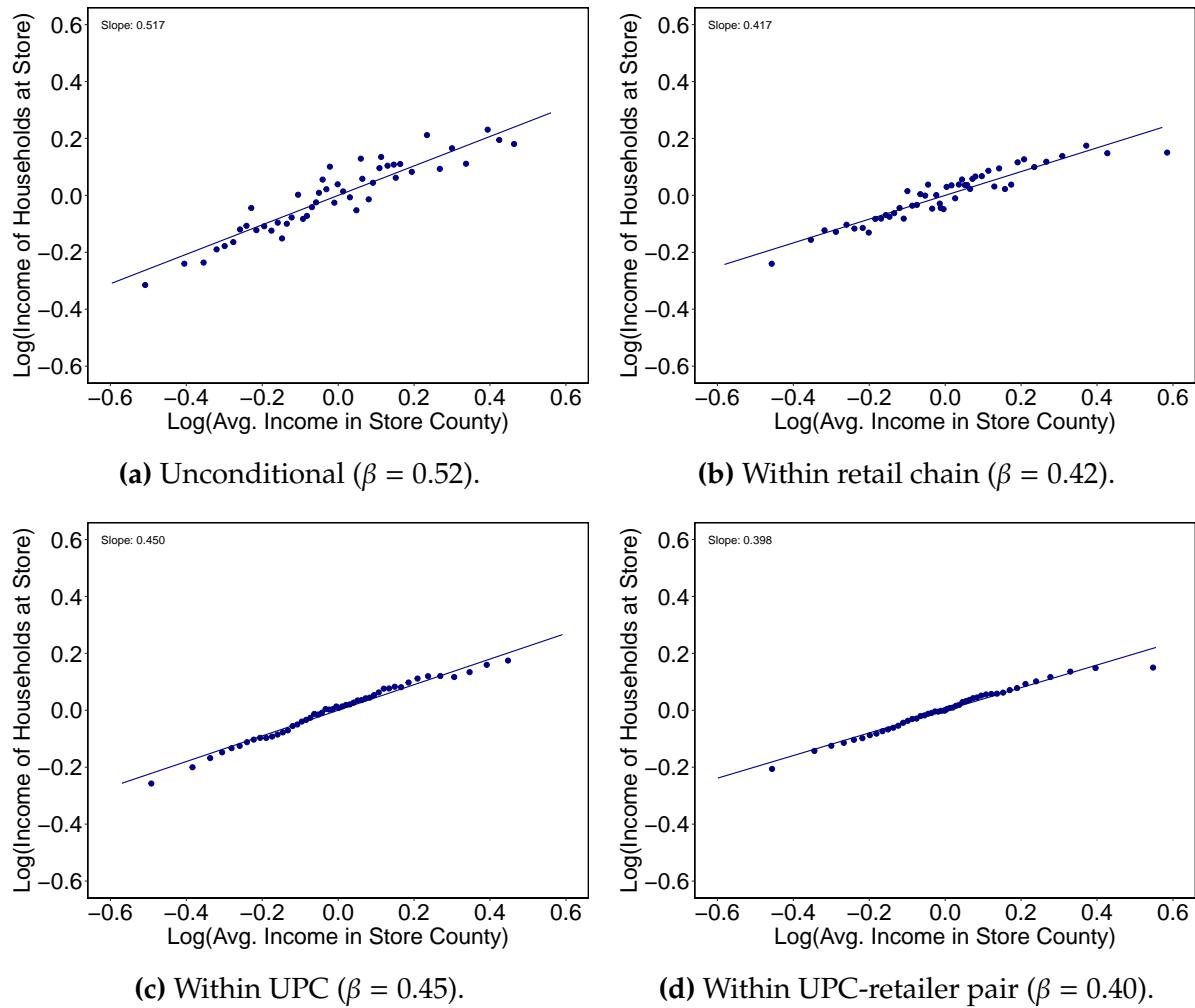
	Retailer-level
Good specific component	35%
Retailer-good specific component	29%
Transaction component	15%
Covar: good, retailer-good	8%
Covar: good, transaction	3%
Covar: retailer-good, transaction	11%
Retailer specific component	35%
Retailer-good specific component	34%
Transaction component	15%
Covar: retailer, retailer-good	2%
Covar: retailer, transaction	3%
Covar: retailer-good, transaction	11%

different retail chains, the demand conditions facing stores within a retail chain across high- and low-income counties do not vary as much as average county income would suggest. In particular, if only high-income individuals across all counties frequent Retail Chain A, while only low-income individuals across all counties frequent Retail Chain B, then it may be that the elasticity of demand facing Retail Chain A (Retail Chain B) may be low (high) regardless of county income.

To measure the degree of self-selection by consumer income into different retail chains, I measure how household income for transactions in the NielsenIQ Homescan panel varies across stores in counties with different income levels unconditionally and within-retailer. If self-selection is large, the relationship between household income and average county income should be much flatter within-retailer than it is unconditionally.

Panels (a) and (b) of Figure F7 shows that the relationship between buyer income and county income is about 20 percent flatter within-retailer than it is unconditionally (an elasticity of 0.42 vs. 0.52). Since retailers exhibit uniform pricing at the UPC level, panels (c) and (d) report results from a similar exercise that conditions on UPC. Again, the within-retailer relationship is only slightly flatter than the unconditional relationship ($1 - 0.40/0.45 \approx 12\%$). Hence, while there is some self-selection by income group into retailers, this may explain only a small part of the tendency for retailers to set uniform prices across locations.

Figure F7: Do income groups self-select into retail chains? Relationship between buyer income and county income, unconditionally and within-retailer.



Appendix G Alternative Model: Sequential Search

As an alternative to nonsequential search model in the main text, in this appendix I develop a model in which households search for products sequentially, building on the labor market search model by Burdett and Mortensen (1998). Relative to Burdett and Mortensen (1998), the two key innovations are that (1) households search intensities are endogenous outcomes of household maximization decisions, rather than exogenous technologies, and (2) these search intensities are heterogeneous across households.

The model speaks to some features of consumer behavior that are abstracted away in the model in the main text. For example, households in this sequential search model form attachments to retailers—which speaks to the fact that, in the data, households often frequent the same store locations—but also receive sporadic information about prices at other retailers.

A calibration of this model suggests that many results from the nonsequential search model in the main text carry over to this model. First, estimated opportunity costs of search effort across households are again increasing and convex in log income. Second, the calibrated model predicts an 11pp rise in the aggregate markup over time, which is similar in magnitude to predictions from the nonsequential search model in the main text. Third, the predicted rise in the aggregate markup over time is driven in part by rising income dispersion after 1980 and by both within-firm increases in markups and reallocations across firms.

G.1 Model Setup

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

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

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 the steady-state distribution of prices paid by a household i first-order stochastically dominates the distribution of 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 outflows of households of type i matched to a retailer with price greater than or equal to p :

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

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 :

$$G_i(p) = \frac{(\delta + \lambda_i)F(p)}{\delta + \lambda_i F(p)}. \quad (28)$$

Proposition 3 (Decrease in search intensity results in FOSD shift in prices paid.). *Consider two household types i and j . If $\lambda_i < \lambda_j$, then $G_i(p)$ first-order stochastically dominates $G_j(p)$.*

Proof. Using (28), the difference between $G_i(p)$ and $G_j(p)$ is given by:

$$G_j(p) - G_i(p) = \frac{\delta F(p)(1 - F(p))}{[\delta + \lambda_i F(p)][\delta + \lambda_j F(p)]} (\lambda_j - \lambda_i), \quad (29)$$

Thus, if $\lambda_i < \lambda_j$, then $G_i(p) \leq G_j(p)$ for all p , with the inequality holding strictly for $F(p) \in (0, 1)$. Thus, G_i first-order stochastically dominates G_j . ■

Equation (29) conveys the main intuition for how cross-sectional differences in search intensity relate to differences in prices paid: households that exert greater search intensity pay lower average prices. We can now use the relationship between search intensity and prices paid to pin down households' endogenous choices of search effort.

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 a search productivity shifter 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:

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

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

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:

$$\phi(z) = \delta(1 - \lambda(z)) \left[\frac{1}{(\delta + \lambda(z))^2} R - \int_p^R \frac{\delta - \lambda(z)F(p)}{[\delta + \lambda(z)F(p)]^3} pdF(p) \right]. \quad (30)$$

Taking the comparative static with respect to z yields Proposition 4.

Proposition 4. *If $\phi(z)$ is increasing in z , then search intensity $\lambda(z)$ is decreasing in z .*

Proposition 4 pins down the relationship between income and search intensity: as long as opportunity costs of search effort are increasing with income, then search intensity declines with income.

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 $p \leq R$ 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 $p \leq R$ 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 $p \notin \text{supp}(F)$ makes profits strictly less than π .

Note that we can rule out distributions F where the maximum price $\bar{p} > R$. This is because, as long as $R > mc$, $\pi(R) > 0$. But for any $p > R$, $\pi(p) = 0$, and so the maximum price in the support of F cannot be strictly greater than 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} :

$$\pi(\bar{p}) = (\bar{p} - mc) \frac{\delta}{M} \int_0^\infty \frac{\lambda(z)}{(\delta + \lambda(z))^2} dH(z). \quad (31)$$

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 :

$$\underline{p} = mc + \delta^2 (R - mc) \frac{\int_0^\infty \frac{\lambda(z)}{(\delta + \lambda(z))^2} dH(z)}{\int_0^\infty \lambda(z) dH(z)} \quad \text{and} \quad \frac{\int_0^\infty \frac{\lambda(z)}{(\delta + \lambda(z))^2} dH(z)}{\int_0^\infty \frac{\lambda(z)}{[\delta + \lambda(z)F(p)]^2} dH(z)} = \frac{p - mc}{R - mc}. \quad (32)$$

Equilibrium. Given R , mc , and household distribution $H(z)$, an equilibrium is a tuple $(\{\lambda(z)\}_{z=0}^\infty, F, \pi, M)$ 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 (30) for all z ; π satisfies (31) with $\bar{p} = R$; M is such that $\pi = f_e$; and $F(p)$ is given by (32) for all $p \in [\underline{p}, R]$, is zero for $p < \underline{p}$ given in (32), and is one for $p > R$.)

In equilibrium, the aggregate markup is

$$\bar{\mu} = 1 + \delta \left(\frac{R}{mc} - 1 \right) \frac{\int_0^\infty \frac{\lambda(z)}{(\delta + \lambda(z))^2} dH(z)}{\int_0^\infty \frac{\lambda(z)}{\delta + \lambda(z)} dH(z)}. \quad (33)$$

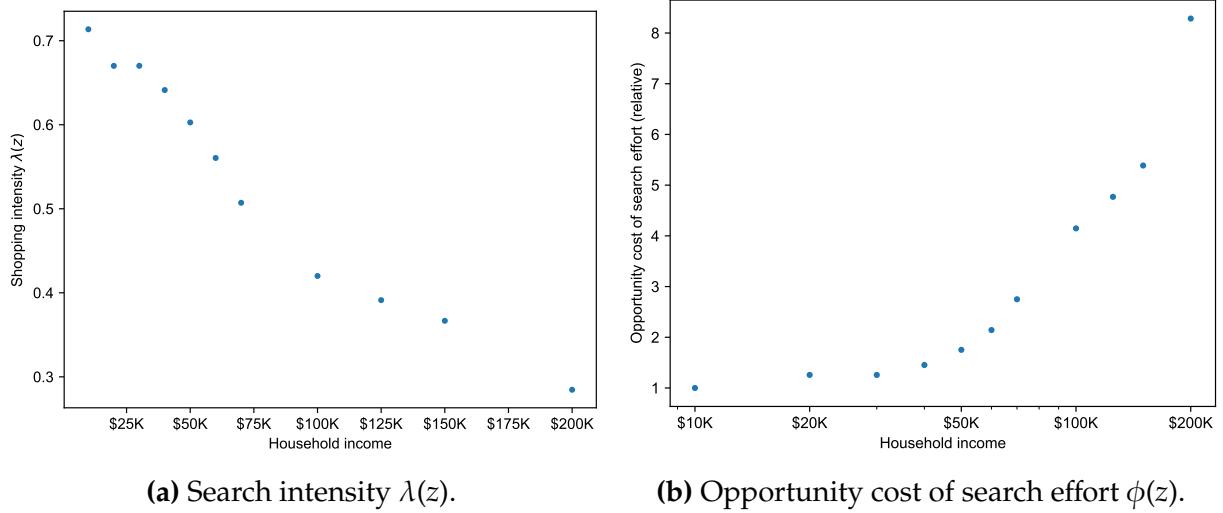
G.2 Calibration and Change in Markups Over Time

As in the main text, I set $R = 3.3$, which is equal to the 98th percentile of retail markups paid in the data. For the rate of exogenous match destructions, I choose $\delta = 0.10$. Note, however, that the choice of δ is not material for the calibration results: as shown in (32) and (33), the distribution of prices and the aggregate markup depend only on the ratio of $\lambda(z)$ to δ , so the choice of δ simply scales the estimated search intensities $\lambda(z)$ by a constant.

To calibrate the remaining parameters of the model, I follow a two-step procedure analogous to the one described in the main text. The first step solves for a fixed point in search intensities $\lambda(z)$ and the distribution of offered prices F to match average markups paid by income group. The second step then solves for opportunity costs of search effort $\phi(z)$ from the household first order condition (30).

Figure G1 shows the calibrated search intensities, $\lambda(z)$, and opportunity costs of search

Figure G1: Calibrated shopping intensity $\lambda(z)$ and search productivity $a(z)$.



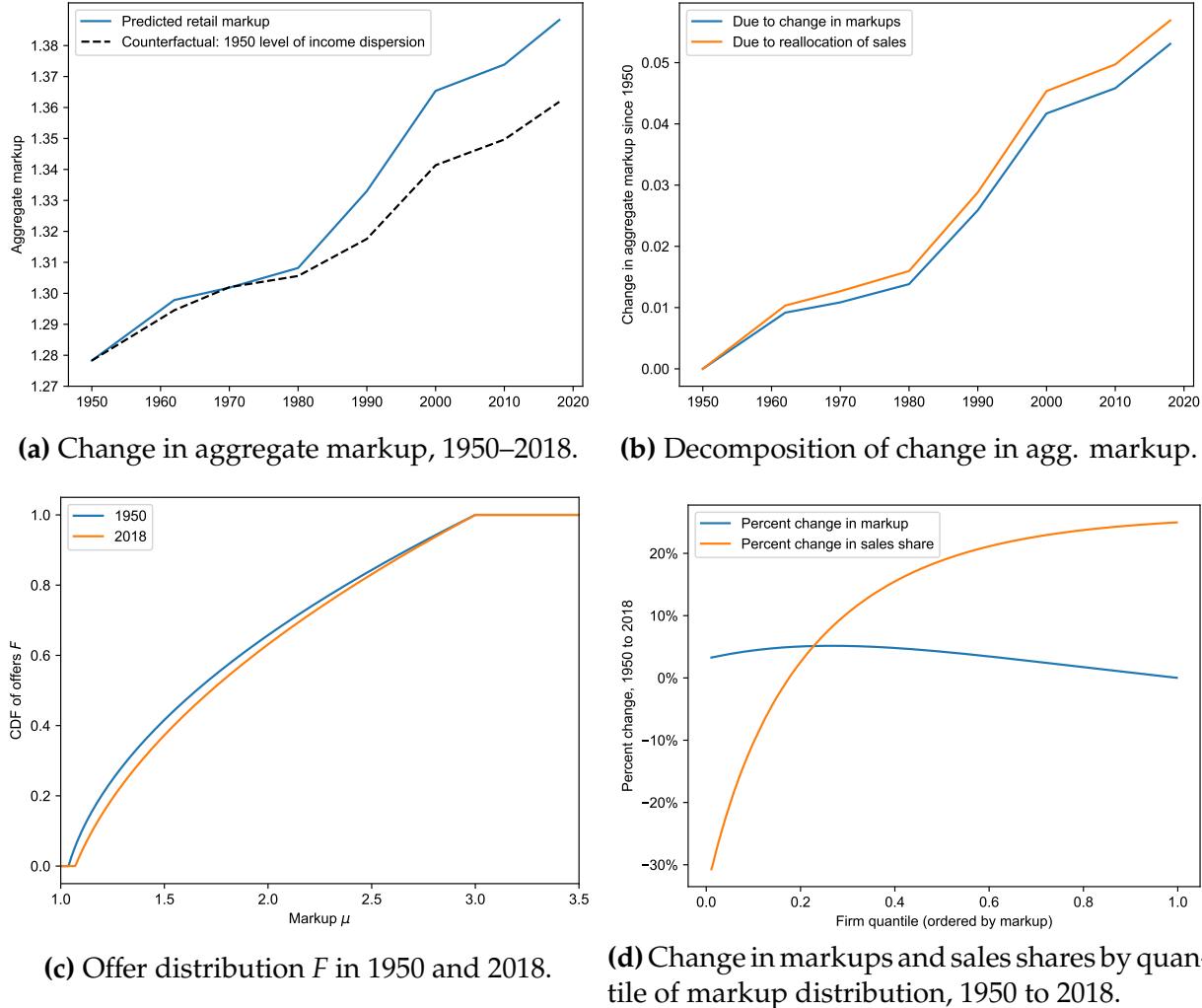
effort, $\phi(z)$, by income group. Search intensities of low-income households are two times greater than those of the highest-income households in the data. Accordingly, opportunity costs of search effort are increasing and convex in income, like those in the nonsequential search model in the main text.

I use the model to simulate how markups change over time as the income distribution changes, holding the mapping between post-tax real income and search intensity fixed. Table G1 reports that changes in the income distribution from 1950 to 2018 account for an 11.0pp increase in the aggregate retail markup through the lens of the model. As in the calibration in the main text, the rise in markups accelerates after 1980 due to both the rising level and dispersion in incomes (Figure G2(a)). Changes in income dispersion from 1980 to 2018 are responsible for 30 percent of the predicted rise in the aggregate retail markup during this period. While changes in the aggregate markup are partly driven by increases in markups at all quantiles of the firm markup distribution, there is also a reallocation of sales to high-markup firms at the top of the markup distribution (Figure G2(b) and (d)). This reallocation of sales to high-markup firms constitutes just over half of the rise in retail markups predicted by the model.

Table G1: Predicted change in aggregate retail markup from 1950–2018.

Period	Predicted Δ in markup	Due to		Due to	
		Δ Income level	Δ Income dispersion	Within-firm changes	Cross-firm reallocations
1950–2018	11.0pp	8.4pp	2.6pp	5.3pp	5.7pp
1950–1980	3.0pp	2.7pp	0.3pp	1.4pp	1.6pp
1980–2018	8.0pp	5.6pp	2.4pp	4.0pp	4.0pp

Figure G2: Sequential search model: Predictions for retail markups from 1950–2018.



Appendix H Alternative Model: Non-Homothetic Preferences

In the main text, differences in price sensitivity across income groups arise endogenously due to differences in search effort. This microfoundation contrasts with a large literature that attributes differences in price sensitivity to non-homothetic preferences or static utility primitives (e.g., Berry et al. 1995; Simonovska 2015; Handbury 2021; Auer et al. 2022).

To contrast these approaches and their implications, this appendix develops a model in which differences in markups across households arise due to non-homothetic preferences. In particular, I follow Handbury (2021) and Faber and Fally (2022) by allowing for non-homotheticities in elasticities of substitution and tastes for quality. As discussed by Handbury (2021), these preferences capture two of the most common forms of non-homotheticities explored in the international literature.

I find that both non-homotheticities in elasticities of substitution and tastes for quality are necessary to account for the patterns in the data. First, large differences in elasticity of substitution are necessary to account for differences in markups paid across income groups. Previous studies that do not include data on markups across the income distribution, such as Handbury (2021) and Faber and Fally (2022), find that non-homotheticities in elasticities of substitution are minor or unnecessary, in contrast. Second, while different elasticities of substitution endogenously generate differences in income groups' expenditure shares across varieties, non-homotheticities in tastes for quality are still necessary to match the extent to which spending patterns by income groups differ in the data.

I contrast the implications for markups across cities and changes in markups over time in this model with those in the main text. In the model developed in this appendix, the elasticity of predicted markups to income across cities is nearly three times larger than in the data. Changes in the income distribution over time increase the aggregate markup from 1950 to 2018 by over 20pp. Both predictions overshoot the predictions of the search model in the main text because the search model features a dampening feedback loop from strategic interactions in search. I interpret the differences between the calibration results as suggesting that modeling differences in price sensitivity with utility primitives may perform poorly out-of-sample. These differences motivate modeling the underlying strategic choices that dictate price sensitivity.

H.1 Model Setup

Households. Households consume a continuum of varieties indexed by $k \in [0, 1]$. Household i 's utility is strictly increasing in its consumption C_i , which satisfies

$$C_i = \left[\int_0^1 [\beta_k(C_i)]^{\frac{1}{\sigma(C_i)}} (c_{ik})^{\frac{\sigma(C_i)-1}{\sigma(C_i)}} M dk \right]^{\frac{\sigma(C_i)}{\sigma(C_i)-1}}, \quad (34)$$

where c_{ik} is household i 's consumption of variety k , M is a measure of the mass of firms, $\beta_k(C_i)$ is a taste shifter for variety k that is allowed to vary with household i 's total consumption C_i , and $\sigma(C_i)$ is an elasticity of substitution across varieties that is also allowed to vary with C_i .

The preferences in (34) are based on preferences used by Handbury (2021) and Faber and Fally (2022), but with two main differences. First, (34) assumes a continuum of measure-zero varieties, thus abstracting away from oligopolistic forces. Given the absence of a positive relationship between market share and retail markups (see Appendix F.1), these oligopolistic forces are not critical to match the patterns in the data. Second, in Handbury (2021), non-homotheticities depend on the households' non-grocery expenditures (which are later proxied for with total household income), while in (34), non-homotheticities instead depend on the size of the total consumption basket. This modeling choice allows me to abstract from the additional complexity of different grocery and non-grocery sectors.

Households each supply one unit of labor inelastically. The labor productivity, and hence wage, of household i is denoted z_i (where the price of one unit of labor is the numeraire). Household i thus chooses how much to consume of each variety in order to maximize utility, subject to the budget constraint,

$$\max_{\{c_{ik}\}} C_i \quad \text{s.t.} \quad \int_0^1 p_k c_{ik} M dk = z_i,$$

where we anticipate that free entry (discussed below) will set the labor share of income to one.

For ease of exposition, denote the expenditure share of household i on variety k by χ_{ik} ,

$$\chi_{ik} = \frac{p_k c_{ik}}{\int_0^1 p_k c_{ik} M dk},$$

and denote the aggregate markup paid by household i as μ_i^{agg} .

Firms. Firms each supply a single variety and produce with a constant-returns technology with a per-unit cost of $1/A_k$. Firms set prices to maximize profits, taking as given all other prices:

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

where aggregate demand C_k for variety k integrates over demand from each household $i \in \mathcal{I}$:

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

By solving the firm's maximization problem, we find that the profit-maximizing price for firm k is equal to its marginal cost times a markup μ_k given by the classic Lerner formula,

$$p_k = \mu_k \frac{1}{A_k} = \frac{\sigma_k}{\sigma_k - 1} \frac{1}{A_k}, \quad \text{where} \quad \sigma_k = \frac{\int_{\mathcal{I}} \sigma(C_i) p_k c_{ik} di}{\int_{\mathcal{I}} p_k c_{ik} di}. \quad (35)$$

Note that the aggregate elasticity of demand facing the firm, σ_k , is equal to the expenditure share-weighted average of the elasticities of substitution of variety k 's consumers.

The mass of firms M is pinned down by free entry. Firms pay a fixed entry cost f_e in units of labor and subsequently realize their type k . The free entry condition requires that expected post-entry profits equal the fixed entry cost,

$$\int_0^1 \left(p_k - \frac{1}{A_k} \right) C_k dk = f_e.$$

Equilibrium. In equilibrium, households choose consumption of each variety to maximize utility, firms choose prices to maximize profits, the zero profit condition holds, and resource constraints are satisfied.

H.2 Calibration

I calibrate the model in two steps. In step 1, I take expenditure shares from the data and choose elasticities of substitution, $\sigma(C_i)$, to match the average retail markups paid by income group. In step 2, I use the expenditures shares from the data and the calibrated elasticities of substitution to back out implied prices and tastes for quality. I discuss each of these steps in more detail.

Step 1: Expenditure shares and elasticities of substitution. For step 1, I begin by taking expenditure shares χ_{ik} directly from the data. In principle, one could define each UPC in

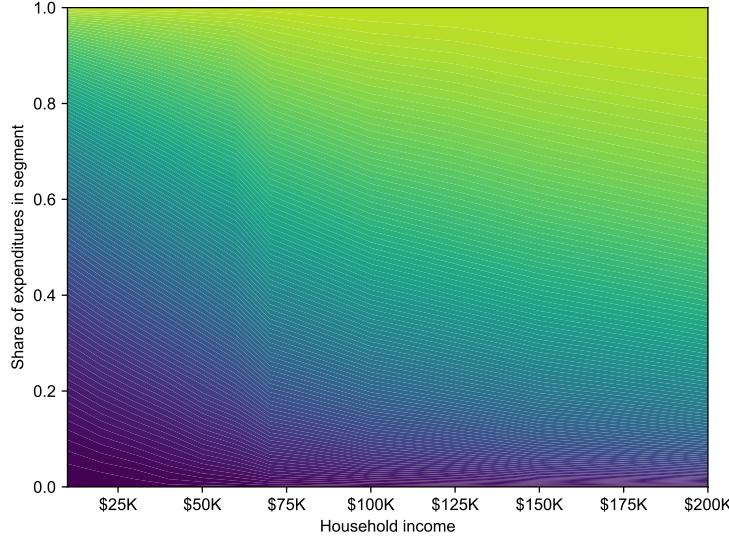


Figure H1: Expenditure shares χ_{ik} , using $K = 100$ segments ordered by average buyer income.

the data as a separate variety. For parsimony—and for a transparent comparison with the search model calibration in the main text—I instead order all UPCs in the data from lowest to highest average buyer income and split the UPCs into K segments with equal sales. Each of the K segments in the data corresponds to a separate variety in the calibration. Figure H1 plots the expenditure shares χ_{ik} of households with incomes z_i on each of the K varieties. As expected, low-income households have higher expenditure shares on lower-ranked segments, while high-income households have higher expenditure shares on higher-ranked segments.

I then choose elasticities of substitution to match the observed aggregate retail markups paid by each income group. To reduce dimensionality, I impose that $\sigma(C_i)$ is a polynomial in log income.⁶² Figure H2 plots the resulting estimates of elasticities of substitution by income. Since aggregate markups paid increase with income, estimated elasticities of demand fall with income from 7 for low-income households to under 2 for the highest-income households in the sample. For comparison, Figure H2 also plots estimates from Auer et al. (2022), who identify differences in substitution patterns of low- and high-income households using a Swiss exchange rate shock. The differences in elasticities of substitution in my calibration line up closely with estimates from Auer et al. (2022).

⁶²For the results presented here, I assume $\sigma(z_i)$ is a third-order polynomial in log income. Specifying a second- or fourth-degree polynomial for this relationship does not significantly affect the results.

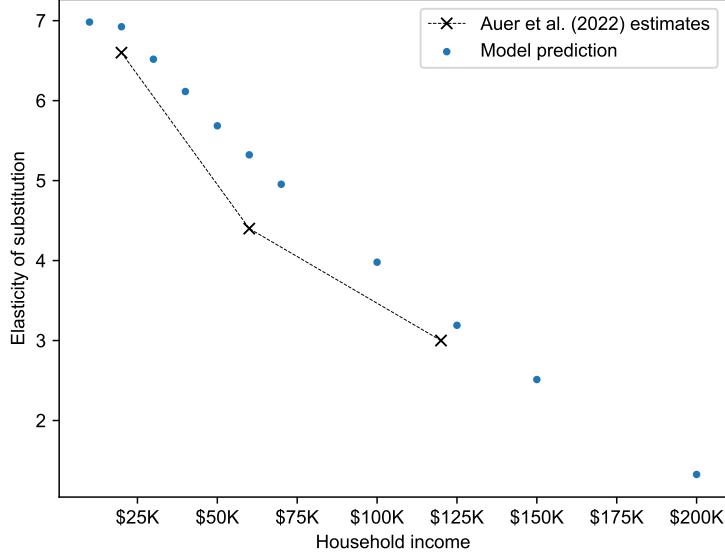


Figure H2: Calibrated elasticities of substitution $\sigma(z_i)$, compared to estimates from Auer et al. (2022).

Step 2: Prices and quality shifters. In step 2, I use the expenditure shares from the data and the estimated elasticities of substitution to back out prices and tastes for quality. I begin with the observation that, for a household i , the ratio of expenditures on varieties k and k' satisfies

$$\log \frac{\chi_{ik}}{\chi_{ik'}} = (1 - \sigma_i) \log \frac{p_k}{p_{k'}} + \log \frac{\beta_{ik}}{\beta_{ik'}}. \quad (36)$$

Differencing (36) for two households i and j with different income levels yields,

$$\log \frac{\chi_{ik}}{\chi_{ik'}} - \log \frac{\chi_{jk}}{\chi_{jk'}} = \underbrace{-(\sigma_i - \sigma_j) \log \frac{p_k}{p_{k'}}}_{\text{Due to different } \sigma} + \underbrace{\left(\log \frac{\beta_{ik}}{\beta_{ik'}} - \log \frac{\beta_{jk}}{\beta_{jk'}} \right)}_{\text{Due to different tastes}}. \quad (37)$$

Equation (37) makes it clear that the relative expenditures on two varieties can differ across income groups for two reasons. First, since income groups have different elasticities of substitution, an income group with higher σ_i will have a lower relative expenditure share of more expensive varieties. This first channel means that, even in the absence of non-homotheticities in tastes for quality, spending patterns will differ across income groups. Second, income groups may also have different tastes for varieties, where a higher β_{ik} will increase the relative expenditure share for household i on variety k .

Since prices and quality shifters are only identified up to log-differences, I normalize quality shifters for the lowest income group to be equal to one for all varieties, quality



Figure H3: Calibrated quality shifters, $\log \beta_{ik}$.

shifters for the first product segment to be equal to one for all income groups, and the price of the first product segment p_0 to be normalized to one. Thus, prices p_k are given by

$$\log p_k = \frac{1}{1 - \sigma_L} \log \frac{\chi_{Lk}}{\chi_{L0}},$$

where L denotes the lowest-income group, $k = 0$ is the first product segment. With prices in place, income group-specific taste shifters are pinned down by (37) and productivities A_k are pinned down by (35).

Figure H3 plots the estimated quality shifters $\log \beta_{ik}$ for each income group and each variety. If non-homotheticities in elasticities of substitution were sufficient to explain the differences in spending patterns by income groups, minimal differences in quality shifters would be needed to explain the data. Instead, we see that large quality shifters are needed to rationalize the high expenditure shares of high-income households on high-ranked varieties. This result means that matching differences in spending patterns across income groups requires a non-homotheticity in taste for quality.

H.3 Markups Across Cities and Changes in Markups Over Time

I begin by using the model to predict markups across cities, using the distribution of incomes across CBSAs from the American Community Survey (ACS) five-year estimates. Table H1 reports that the model predicts aggregate retail markups across cities rise with

Table H1: Non-homothetic preferences model predictions for markups across cities, compared to the data.

<i>Log CBSA Markup</i>	Model-Predicted		Data	
	(1)	(2)	(3)	(4)
Log CBSA Income	0.297** (0.005)	0.278** (0.005)	0.110** (0.006)	0.102** (0.007)
Gini Index		0.369** (0.042)		0.153** (0.057)
<i>N</i>	881	881	881	881
<i>R</i> ²	0.82	0.84	0.27	0.28

Note: The dependent variable in columns 1–2 is the aggregate markup for the CBSA predicted by the nonhomothetic preferences model, and in columns 3–4 is the cost-weighted average markup across all CBSA purchases in the data. Regressions weighted by total CBSA sales in the data. ** is significant at 5%.

both per-capita income and inequality, consistent with retail markups in the data. However, the “macro elasticities” of markups to income and inequality are substantially larger than in the data. For example, doubling per-capita income in a city is predicted to increase markups nearly 30 percent, compared to 11 percent in the data.

To calculate how changes in the income distribution over time affect the aggregate retail markup, I assume that the mapping from post-tax real income to elasticities of substitution and tastes for quality remains fixed over time.⁶³ As in the main text, I use the distribution of post-tax real income documented by Saez and Zucman (2019). I assume that productivities A_k across varieties stay fixed, but that markups μ_k and thus prices p_k update each year as the income distribution changes. Since elasticities of demand depend on the share of expenditures for each product coming from each income group, and income groups’ expenditure shares across products depend on relative prices of products, computing markups under each income distribution requires solving a fixed point.

Table H2 reports that this model predicts a 21.6pp rise in the aggregate retail markup from 1950 to 2018 due to changes in the income distribution. Rising income dispersion contributes 8.8pp, or about 40 percent, of the total rise in markups over this period, and most of the rise in the aggregate retail markup predicted by the model occurs after 1980.

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, overshooting the rise in retail markups in the data substantially. Both quantitative exercises predicting

⁶³This assumption deviates slightly from the preferences in (34), since (34) specifies the mapping in terms of real consumption rather than income. Quantitatively, specifying the mapping in terms of real consumption or income has little effect on the counterfactual results.

Table H2: 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	21.6	12.8	8.8
1950–1980	3.8	3.5	0.3
1980–2018	17.7	9.2	8.5

markups across cities and over time indicate that, while the model is able to match the “micro elasticity” of markups to income in the cross-section, it overshoots the “macro elasticity” of markups to aggregate income.

The key difference between this model and the search model in the main text is that the search model has an additional feedback loop from strategic search decisions that is absent in this model. In the search model, when the composition of households shifts toward higher-income households, aggregate search intensity falls, but households each individually exert more search effort to offset this change. In the model developed in this appendix, households’ elasticities of substitution are static, and so there is no offsetting forces to dampen the fall in aggregate price sensitivity as the economy shifts towards higher income groups.

I interpret these results as warning that using utility primitives to model differences in price sensitivity may perform poorly out-of-sample. In particular, the absence of an interaction between aggregate price sensitivity and individual price sensitivity—which arises endogenously in the search model—results in counterfactual predictions.

Appendix I Spatial Spillovers of Regional Income Shocks

In this appendix, I show that the search model developed in the main text can be used to simulate how markups across areas respond to regional income changes. These exercises are meant to be a proof-of-concept for spatial spillovers of regional income shocks through the channel of demand composition. Richer spatial models can incorporate this spillover channel alongside other interactions.

The key assumption I make for these exercises is that retail chains set uniform prices for goods across locations. This phenomenon is documented by DellaVigna and Gentzkow (2019) and Hitsch et al. (2021). As discussed by DellaVigna and Gentzkow (2019), uni-

form pricing within retail chains could be due to brand concerns or managerial inertia. Regardless of micro-foundation, I assume that retail chains observe the distribution of customers that shop for each good, but are unable to set different prices for the same good at different store locations. As a result, stores in the retail chain randomly vary prices over a distribution of prices described by the search model in the main text.⁶⁴

Uniform pricing across store locations results in spatial spillovers because an increase in incomes in one area may lead a retail chain to increase markups across all locations. Note that the empirical evidence on spillovers in Section 3.2 exploits exactly this source of variation in incomes of customers at a retail chain's other locations and documents positive spillovers on markups paid.

I consider two counterfactual exercises in this appendix. First, I consider how markups change across CBSAs if we permanently double incomes of all households initially earning over \$100K in the New York - Newark - Jersey City MSA. Second, I consider how markups change if we permanently halve the incomes of all households initially earning over \$100K in the two CBSAs that constitute the San Francisco Bay Area.

I begin by constructing a dataset of expenditures for $K = 10$ groups of goods at each retail chain by CBSA and income group. (I use the same procedure described in the main text to construct the 10 groups.) I use the values of $\phi(z)$ calibrated in the main text to predict initial markups that each retail chain sets for each of the K goods.

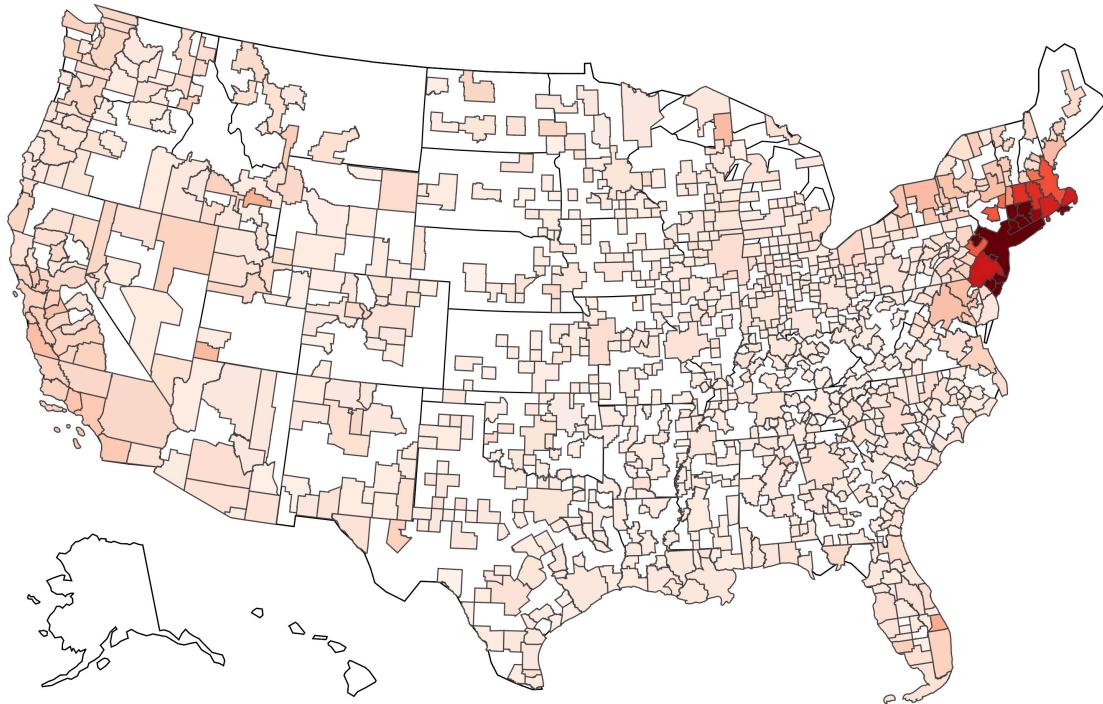
For each counterfactual, I then re-predict markups across all K goods at all retail chains after including a regional income shock. I make the following simplifying assumptions: (1) after an income shock, a group takes on the opportunity cost of search effort $\phi(z)$ given by its new income value; (2) total expenditures by the group also scale according to the change in the group's income; (3) all income groups maintain identical spending shares across retailers as in the initial calibration. These assumptions could easily be relaxed in a richer model. Finally, for both counterfactuals, I aggregate the initial and counterfactual markups to get changes in markups by CBSA.

Figure I1 shows the results for the first counterfactual in which high-end incomes in New York are doubled. There is a large increase (5.5pp) in markups paid in the New York City MSA, due both to the shift in composition of expenditures to higher-income households and the effects of firms increasing their markups in response to lower aggregate search intensity. A number of other CBSAs in New Jersey and Connecticut see increases in aggregate markups over 1pp, due to spillovers from uniform pricing. The

⁶⁴To maintain the connection to the model in the main text, I abstract away from the fact that a retail chain with multiple stores can control multiple prices and thus may not need to take the distribution of other stores' prices as given.

vast majority of other CBSAs see minimal changes in markups (less than 0.5pp).

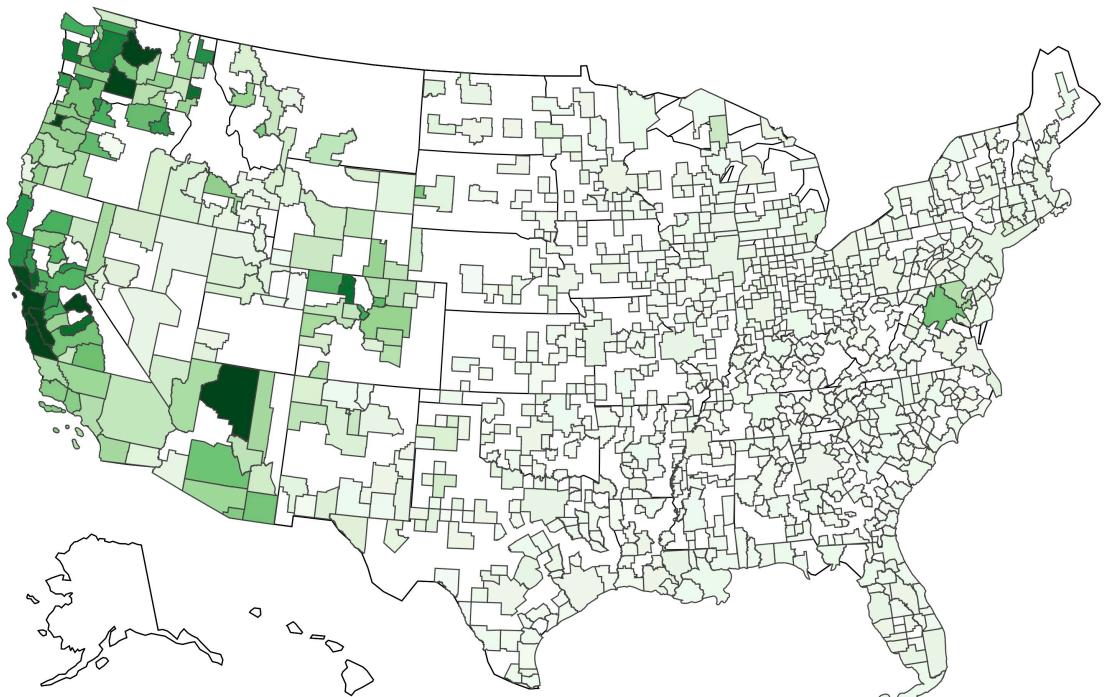
Figure I1: Counterfactual 1: Spillovers from doubling high-end incomes in New York.



CBSA	Change in agg. markup
New York - Newark - Jersey City, NY-NJ-PA	+5.49pp
Atlantic City - Hammonton, NJ	+1.80pp
Bridgeport - Stamford - Norwalk, CT	+1.74pp
Ocean City, NJ	+1.68pp
Trenton - Princeton, NJ	+1.47pp

Similarly, Figure I2 shows results from the second counterfactual in which high-end incomes in the San Francisco Bay Area are halved. Markups in both CBSAs in the Bay Area fall by about 6pp, reflecting both compositional changes in expenditures and firms' responses to the income shocks. There are also spillovers across nearby CBSAs in California, Washington, Oregon, and Arizona.

Figure I2: Counterfactual 2: Spillovers from halving high-end incomes in the Bay Area.



CBSA	Change in agg. markup
San Jose - Sunnyvale - Santa Clara, CA	-6.19pp
San Francisco - Oakland - Berkeley, CA	-5.90pp
Santa Cruz - Watsonville, CA	-1.20pp
Yakima, WA	-1.18pp
Sonora, CA	-1.07pp

References for Online Appendix

- 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.
- 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.
- Baker, S. R., B. Baugh, and M. C. Sammon (2020). Measuring customer churn and interconnectedness. 27707, National Bureau of Economic Research.
- Barger, H. (1955). *Distribution's Place in the American Economy Since 1869*, Chapter "Trends in Margins", pp. 80–90. NBER.
- 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.
- Berry, S., M. Gaynor, and F. S. Morton (2019). Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives* 33(3), 44–68.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4), 841–890.
- Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics* 121, 1–14.
- 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.
- 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 D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.
- Butters, R. A., D. W. Sacks, and B. Seo (2022). How do national firms respond to local cost shocks? *American Economic Review* 112(5), 1737–1772.
- Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *The Journal of Finance* 63(4), 1977–2011.
- Conlon, C. and J. Gortmaker (2020). Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics* 51(4), 1108–1161.
- 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.

- 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.
- Griffith, R., L. Nesheim, and M. O'Connell (2010). Sin taxes in differentiated product oligopoly: An application to the butter and margarine market. Technical Report CWP37/10, CEMMAP Working Paper.
- 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.
- Hitsch, G. J., A. Hortaçsu, and X. Lin (2021). Prices and promotions in us retail markets: Evidence from big data. *Quantitative Marketing and Economics*, 1–80.
- Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics* 131(3), 1291–1364.
- 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.
- Kim, D. (2008). Demand and pricing in the us margarine industry. *Journal of Agricultural and Food Industrial Organization* 6(1).
- Lach, S. (2002). Existence and persistence of price dispersion: An empirical analysis. *Review of Economics and Statistics* 84(3), 433–444.
- 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.
- 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.
- 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.
- 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. and M. Price (2018). The new gilded age: 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.
- Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.
- Villas-Boas, S. B. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *Review of Economic Studies* 74(2), 625–652.

Wu, L. (2022). Network lerner index: Demand and distortions across industries. Working paper.