Markups Across the Income Distribution: Measurement and Implications

Kunal Sangani
Stanford University*

October 31, 2024

Abstract

We examine the relationship between customer income and firm markups using data on household transactions and wholesale costs. Doubling a household's income is associated with a 2–3 percent increase in retail markups paid. This elasticity of markups to household income is two times larger than previous estimates that compare prices paid for identical products alone. We estimate that doubling the income of all households in an economy leads to an 8–15 percent increase in the aggregate markup due to spillovers across households. We develop a macroeconomic model with consumer search that can account for these facts. Consistent with the model's predictions, we 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 11pp rise in retail markups, with 25 percent of the increase due to growing income dispersion. This increase consists of both within-firm markup increases and a reallocation of sales to high-markup firms, which occurs without any change to the nature of firm production or competition.

Keywords: Markups, search, income distribution.

**Email:* 8ksangani@gmail.com. First version online: March 2022. I am grateful to Gabriel Chodorow-Reich, Andrei Shleifer, and Ludwig Straub for their advice and support. This paper also benefited from comments by David Baqaee, Adrien Bilal, Xavier Gabaix, Jeff Gortmaker, Erik Hurst, Alexander MacKay, Guido Menzio, 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 (e.g., Autor et al. 2020; Barkai 2020; De Loecker et al. 2020). 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 the role of the income distribution in shaping markups across space and time. We make three contributions. First, using rich data on household purchases and wholesale prices, we measure how retail markups vary with an individual household's income—which we term the *micro elasticity* of markups to income—and how retail markups in an economy vary with aggregate income—the *macro elasticity* of markups to income. Second, we develop a macroeconomic model of consumer search that can account for these empirical patterns. Third, using both model-free and model-based estimates, we find that changes in the income distribution over time can plausibly account for a substantial portion of the rise in retail markups observed in the data.

We construct a dataset of 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 for the exact items (product barcodes) purchased by households. Our 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 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 26 million transactions made in a single year.

[1] Exceptions 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.

[2] Two other ways markups are commonly measured are demand estimation and production function estimation. We show in Appendix F that markups recovered using both alternative approaches also exhibit similar patterns to the retail markups we use in the main text.

We use these data to measure the relationship between markups and household income. First, we measure how markups paid vary with an individual household's income, which we refer to as the "micro elasticity" of markups to income. The data indicate stark differences in markups paid across income groups. Our most conservative estimates, which control for unobserved local costs by comparing markups paid by households shopping at the same store outlet, indicate that doubling a household's income is associated with a 2.2 percent increase in retail markups paid. Allowing for comparisons in markups paid across stores increases the micro elasticity of markups to income to 3.4 percent. We conduct several analyses to check that this relationship between markups and household income is robust to unobserved cost shifters such as local input costs, shipping and transport costs, spoilage costs, and retailer-specific negotiated discounts.

The micro elasticity of markups to income is two times larger than previous estimates that compare prices paid for identical products across income groups (e.g., Aguiar and Hurst 2007a; Broda et al. 2009). The elasticity we document is larger because it accounts for differences in markups across products. In addition to paying higher prices for identical products, high-income households purchase a basket of products that have higher markups on average. The use of cost data allows us to identify how basket composition contributes to differences in markups paid across income groups for the first time.[3]

While these estimates describe how markups vary with an individual household's income, how markups vary with aggregate income—i.e., the "macro elasticity" of markups to income—also depends on spillovers that are absorbed in the intercept of cross-sectional regressions (Chodorow-Reich 2020; Wolf 2023). The sign and magnitude of these spillovers is theoretically ambiguous. To isolate these spillovers, we develop three specifications, each of which exploits a different source of variation in the incomes of other customers that a household shops alongside. For example, we isolate how markups paid by a household change with variation in the income of a retail chain's or product's customers over time, controlling for both time-varying unobserved household characteristics and local costs.

Across the three specifications, we find consistent evidence of positive and large spillovers of other customers' incomes on markups paid. These spillovers imply that the macro elasticity of markups to income is larger than the 2–3 percent micro elasticity observed across households. We estimate that doubling the real incomes of all households in the economy would increase markups 8–15 percent.

[3]Measuring how differences in markups across products contribute to the markup gap is also useful given mixed theoretical predictions: classic quality discrimination models (e.g., Mussa and Rosen 1978) predict that firms charge lower markups on products bought by high-income customers to deter customers from switching to lower quality products, while models with non-homothetic preferences predict higher markups on these products due to the low price sensitivity of high-income buyers.

We use these moments to develop a macroeconomic model in which firms' markups depend on the income distribution. The model features information frictions that lead households to retrieve price quotes from firms, generating dispersed prices for identical products as in the canonical search model by Burdett and Judd (1983). We add two ingredients that cause the income distribution to affect firms' markups. First, we allow the number of price quotes retrieved by a household to depend on its endogenous choice of search effort. Second, we allow households to differ in their tastes for goods—leading to differences in basket composition—and in their labor and search productivities, leading to heterogeneous opportunity costs of time and thus search intensities. As a result, the model generates each of the patterns we document in the data: (1) differences in markups paid for identical products across households, (2) differences in average markups across products, and (3) spillovers across households that lead the macro elasticity to differ from the micro elasticity.

The model is flexible enough to accommodate various patterns between markups and income, but tractable enough for us to characterize how shifts in the income distribution affect markups. For example, we show that depending on how labor and search productivity vary with households' incomes, the model can either generate a "poverty premium" in which the lowest-income households pay the highest markups (e.g., Caplovitz 1963; Prahalad and Hammond 2002) or explain the patterns we observe where markups paid increase with income. We analytically characterize the conditions under which a first-order stochastic shift or a mean-preserving spread in the distribution of buyers' incomes leads to a rise in the aggregate markup. It turns out that both sets of conditions hold when we calibrate the model to fit differences in basket composition and markups paid across income groups in the data.

We use the model to return to the question that opens this paper: how do changes in the income distribution shape markups across space and time? In keeping with our theoretical results, the calibrated model predicts that markups across U.S. cities increase with per-capita income and inequality. Both predictions are borne out in our data on retail markups across U.S. cities. Moreover, our model has much greater success in predicting markups across cities than standard macroeconomic models that predict markups using data on firms' market shares and concentration (e.g., Atkeson and Burstein 2008). We show that markups predicted by such "supply-side" models in fact have a negative correlation with markups across cities in our data, and that the predictions of our demand-side model outperform both those models' predictions and income measures alone in a horse race.

We also find that changes in the income distribution can play a meaningful role in the evolution of markups over time. Through the lens of the model, changes in the distribution of post-tax, real income documented by Saez and Zucman (2019) account for a 11pp rise in the aggregate retail markup from 1950 to 2018. Increases in the aggregate markup are moderate from 1950–1980 but accelerate after 1980 due to rising income dispersion. The path of markups predicted by the model is in line with data on retail gross margins from the Census of Annual Retail Trade Survey.

In the model, nearly half of the rise in the aggregate markup over time is due to a reallocation of sales to high-markup firms. High-markup firms expand because falling search intensity and changes in demand composition lead households to shop more often at firms with high markups. The importance of reallocations is consistent with evidence from Autor et al. (2020) and De Loecker et al. (2020). Rather than the growth of high-markup firms causing the rise in markups, however, in our model both reallocations and rising markups are endogenous consequences of changes to 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.

Related literature. Our estimates of the micro elasticity of markups to income add to a literature on differences in prices paid for identical products across households, including seminal papers by Aguiar and Hurst (2007a), Broda et al. (2009), and Kaplan and Menzio (2015). These studies find that high-income households tend to pay higher prices for identical products, but generally do not have common units for comparing products of different qualities or purchases from stores with different amenities. Our use of cost data to construct retail markups facilitates such comparisons. We find that accounting for cross-product differences in markups doubles the elasticity of markups to household income. These cross-product comparisons are important for characterizing changes in the aggregate markup, since previous research finds that reallocations across products and firms play a key role in the evolution of markups over time (e.g., Autor et al. 2020; Baqaee and Farhi 2020; De Loecker et al. 2020).

We also add to a rich literature in macroeconomics and trade that provides evidence of a relationship between income or wealth, price sensitivity, and markups, including Lach (2007), Alessandria and Kaboski (2011), Simonovska (2015), Anderson et al. (2018), DellaVigna and Gentzkow (2019), Jaimovich et al. (2019), Stroebel and Vavra (2019), Gupta (2020), Handbury (2021), Auer et al. (2022), and Faber and Fally (2022). Most closely related to our macro elasticity of markups to income are Stroebel and Vavra (2019), who estimate an elasticity of retail prices to housing wealth of 15–20 percent, and Anderson et al. (2018), who estimate elasticities of gross margins to income of 10 and 17 percent across stores within two retail chains in Canada and the U.S. The dataset we construct allows us to capture how expenditure patterns across retailers shape the aggregate markup, and to decompose the relationship between markups and income into partial and general equilibrium effects.

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. (2024) find that high-income consumers are less price sensitive in markets for breakfast cereal, yogurt, coffee, and automobiles. The demand systems used in these studies are not well suited to our purposes, however, since they attribute differences in price sensitivity to preferences alone, and thus are unable to account for the fact that differences in prices paid for identical products (even in the same store) are partly responsible for 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.

Our model builds on insights from Aguiar and Hurst (2007a) and Kaplan and Menzio (2015) that cost of time affects shopping behavior and prices paid.[4] Pytka (2018) and Nord (2022) also integrate household search decisions into the Burdett and Judd (1983) model. Both study how search choices interact with incomplete markets to explore the difference between expenditure and consumption inequality.[5] Our model exhibits some key differences in micro-foundation—for example, markups are shaped by a race between labor and search productivity, rather than a disutility for shopping as in Pytka (2018) and Nord (2022)—and allows for any non-parametric distribution of household incomes, enabling analytic comparative statics with respect to the income distribution. The applications to markups across space and over time are also unique to this paper.

Finally, this paper relates to recent work that documents trends in markups over time and considers potential drivers. Brand (2021) and Döpper et al. (2021) estimate demand systems in retail scanner data, and both studies find that demand-side forces play an important role in the evolution of markups they estimate. Brand (2021) attributes rising markups to consumers becoming less price sensitive, perhaps due to increased product differentiation. Döpper et al. (2021) attribute rising markups in scanner data to declining consumer price sensitivity and incomplete pass-through of marginal cost reductions.[6]

[4]Related empirical work on search effort and savings technologies includes Griffith et al. (2009), Aguiar et al. (2013), Coibion et al. (2015), and Nevo and Wong (2019).

[5]Nord (2022), which was released after the first version of this paper, also shows that the skewness of price distributions covaries with buyer income, consistent with a search model. He focuses on prices over the business cycle, while we focus on long-run determinants of markups across space / over time.

[6]Döpper et al. (2021) find that changes in buyer income are partly responsible for markup trends, but conclude that a secular decline in price sensitivity plays an even larger role. This decline in price sensitivity

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

Layout. Section 2 describes the data and our measure of retail markups. Section 3 estimates the micro and macro elasticities of markups to income. Section 4 develops a search model that can account for the patterns in the data, and Section 5 calibrates the model. Section 6 uses the model to explore spillovers across income groups, effects of inequality, and markups across U.S. cities. Section 7 considers how changes in the income distribution over time affect retail markups. Section 8 describes extensions developed in the Online Appendix, and Section 9 concludes.

2 Constructing a Dataset on Retail Markups

This section describes two data sources on consumer purchases and wholesale costs that we use to construct a dataset of retail markups paid by households. Appendix A details how the data are cleaned and describes other ancillary data sources.

2.1 Data sources

Consumer panel data. We use data from NielsenIQ Homescan, which collects data on purchases made by a nationally representative group of households. Homescan panelists use in-home scanners or a mobile application to record purchases of items in Nielsen-IQ tracked categories from any retail outlet. In addition to reporting the date and store visited for 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. NielsenIQ also collects detailed demographic information about participating households on an annual basis, including the household's total income.

The categories tracked by NielsenIQ include grocery products; beauty, personal care, and home care products; and some general merchandise categories such as batteries, candles, and cookware. Broda and Parker (2014) show that the categories tracked by NielsenIQ constitute about 35 percent of households' expenditures on nondurable goods in the consumer expenditure survey and about 19 percent of total consumption. For all

could be due to changes to the income distribution that are not picked up by noisy measures of average buyer income. Our model also provides a rationale for why search behavior conditional on income may shift over time as a strategic response to a changing income distribution.

analyses, we exclude NielsenIQ "magnet products," which include fruits, vegetables, and in-store baked goods that do not use standard UPC codes.

We present results using data from a single year, 2007, which covers 62 million transactions by over 60,000 households. We choose 2007 as our baseline year for analysis since, from 2006 to 2009, NielsenIQ disaggregates households with \$100K, \$125K, \$150K, and over \$200K in income. (In other years, incomes are instead top-coded at \$100K.) The data from 2007 allow us to observe how markups paid vary of this range of incomes, while avoiding the impact of the recession in 2008–2009. For our analysis of spillovers in Section 3.2, we extend our dataset to include purchases from 2006–2012. Following Handbury (2021), we exclude households with below \$10K in income from our analysis.

Wholesale costs. We use data on wholesale costs from PromoData Price-Trak, a service that tracks wholesale prices on a weekly basis for over 100,000 UPCs. The PromoData come from twelve grocery wholesalers that sell goods to retailers across the U.S. Every week, wholesalers send PromoData the 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. In the main text, we present results using deal prices as the measure of retailers' wholesale costs (results from instead using base prices are reported in Table 1).

As previously shown by Stroebel and Vavra (2019), wholesale prices from the PromoData are similar across markets: in Appendix Table A1, we show that 80 percent of market-item pairs in the data have a wholesale price exactly equal to the modal price across markets in that month.[7] Hence, for our baseline results, we calculate a national wholesale price for each UPC in each month. Similar results obtain using the subset of transactions where a wholesale price is reported for the market of purchase (see Table 1).

We match about 67,000 UPCs purchased by Homescan panelists in 2007 to wholesale costs from the PromoData. These UPCs constitute 43 percent of transactions and 37 percent of expenditures in the 2007 panel. Appendix Table A3 shows that the share of transactions and expenditures matched to wholesale costs is similar across income groups.

[7]Using data from a major grocer, DellaVigna and Gentzkow (2019) also find that wholesale costs do not vary across stores (see DellaVigna and Gentzkow 2019 Appendix Figure 14). In Appendix G.1, we use data on prices and costs from a retailer and confirm that differences in markups, rather than differences in costs, account for the vast majority of price differences for a UPC across stores.

2.2 Constructing retail markup estimates

We calculate the retail markup on product (UPC) g purchased by household i in transaction t as the price the household pays over the wholesale cost of product g in the month of transaction t,

$$\text{Retail Markup}_{igt} = \frac{\text{Price}_{igt}}{\text{Wholesale cost}_{gt}}.$$

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 are a reasonable measure of marginal costs since other components of retailers' costs, such as rent, capital, and labor, are fixed at short horizons.

We winsorize retail markups at the 1 percent level for all analyses. The cost-weighted average markup in the dataset is 32 percent, in line with estimates from retailer data and with retail grocery gross margins from the 2007 Census Annual Retail Trade Survey.

Potential sources of measurement error. Our approach of using cost data to measure retail markups follows Eichenbaum et al. (2011), Gopinath et al. (2011), and Anderson et al. (2018), among others. Each of the aforementioned papers uses data provided by a single retailer, and thus observes the accounting measure of replacement costs used internally by the retailer. Relative to using data from a single retailer, the merged dataset that we construct has two advantages for our purposes: (1) it includes all household expenditures in tracked product categories and thus captures how households allocate expenditures across retailers; and (2) it includes detailed demographic information on the households making each transaction, typically not available in retailer scanner data. Observing household demographics and the composition of expenditures across retailers are both crucial for measuring the micro and macro elasticities of markups to income.

These features come with a tradeoff in terms of the granularity of the cost data. First, the list wholesale prices we use do not capture retailer-wholesaler or retailer-manufacturer deals, such as volume discounts, slotting fees, or trade spend, that may cause retailers' replacement costs to differ from the list prices in the PromoData. Second, retailers' true marginal costs may also differ from replacement costs due to local inputs for shelving and inventory management. (Retailers' internal cost measures used in previous work also suffer from some of these biases: internal cost measures tend to omit local input costs, and Anderson and Fox (2019) document inconsistent practices in accounting for rebates and trade spend even within retail chains.)

For these reasons, all our estimates for the micro and macro elasticities of markups to income below control for differences in local input costs (at the county level or at the store outlet level where possible) and for factors that may lead to variation in costs across retail chains.[8] We also compare markups in our dataset to estimates recovered using techniques from the industrial organization literature (see below) and conduct a variety of robustness checks to ensure our results are not driven by mismeasurement of costs.

Comparison to other markup measures. Two other approaches used to measure markups are demand estimation and production function estimation. In Appendix F, we estimate a random coefficients model à la Berry et al. (1995) to recover marginal costs and markups in a product category (margarine) where the PromoData has high coverage. The estimated markups exhibit a strong positive correlation ($\rho \approx 0.6$) with the retail markups in our data.[9] Retailers in the NielsenIQ data are anonymized, so it is more challenging to directly compare our retail markups with firm-level markups measured using the production function approach, but Appendix F shows that markups of public firms measured by De Loecker et al. (2020) exhibit an even stronger positive relationship with customer income than markups in our data.

3 Empirical Findings

In this section, we explore the relationship between markups and household income. Section 3.1 measures the *micro elasticity* of markups to income—how markups vary with an individual household's income. Section 3.2 measures spillovers across households and aggregates our estimates to calculate the *macro elasticity* of markups to income.

3.1 The Micro Elasticity of Markups to Household Income

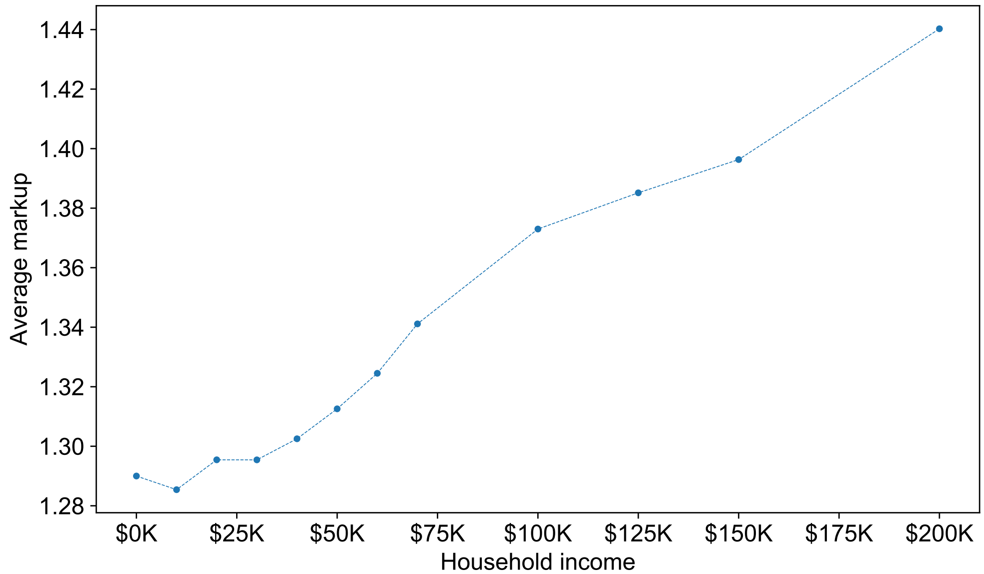
Figure 1 plots the aggregate (cost-weighted average) markup paid by households over the income distribution for the sample of purchases matched to wholesale costs. The aggregate markup increases from 29 percent for the lowest-income households in the sample to 44 percent for households with over $200,000 in annual income.

These descriptive statistics indicate substantial differences in markups paid across income groups. However, these measures of markups may in part reflect unobserved cost shifters, such as local input costs, that we want to control for when measuring the elasticity of markups to income. We first present our most conservative estimates of the

[8]We are helped by the Robinson–Patman Act, which limits the extent to which wholesalers and manufacturers can sell identical products to different retailers at different prices. See our discussion in Footnote 11.

[9]Whether markups recovered from demand estimation include retailers' markups depends on assumptions about vertical conduct between retailers and manufacturers, which we discuss in Appendix F.

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



micro elasticity of markups to income, which control for the store of purchase, and then present our preferred estimates which allow for markup comparisons across stores.

Our most conservative measure of the relationship between markups and household income includes store fixed effects to control for potential cost shifters across stores. For 14 million transactions in our sample, NielsenIQ provides store IDs that identify the specific store outlet where each purchase was made. Using this subsample with store IDs, we estimate markup differences across income groups using the specification,

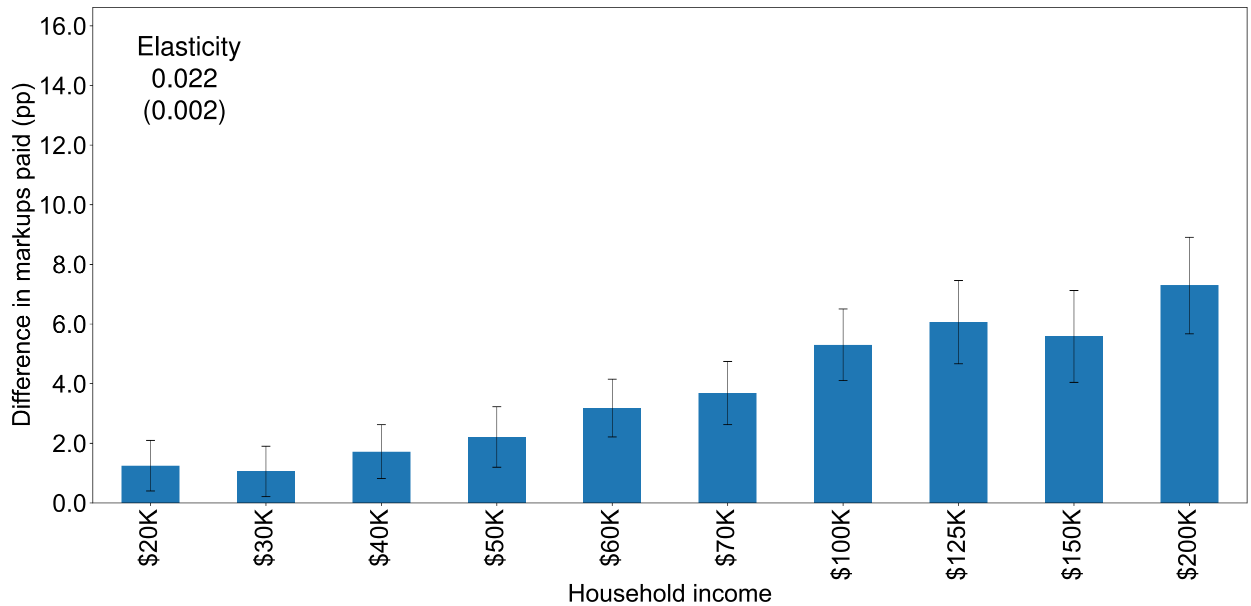
$$\text{Markup}_{ik} = \sum_{\ell} \beta_\ell 1\{i \text{ has income level } \ell\} + \gamma' X_i + \alpha_{s(k)} + \epsilon_{ik}, \tag{1}$$

where Markup_{ik} is the markup paid by household i in transaction k; X_i are household demographic controls that 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;[10] $\alpha_{s(k)}$ are store fixed effects; and ϵ_{ik} is a mean-zero error. We weight the regression by costs and leave out the indicator for households with less than $20,000 in income, so that the coefficients β_ℓ measure differences relative to the group with below $20,000 in income.

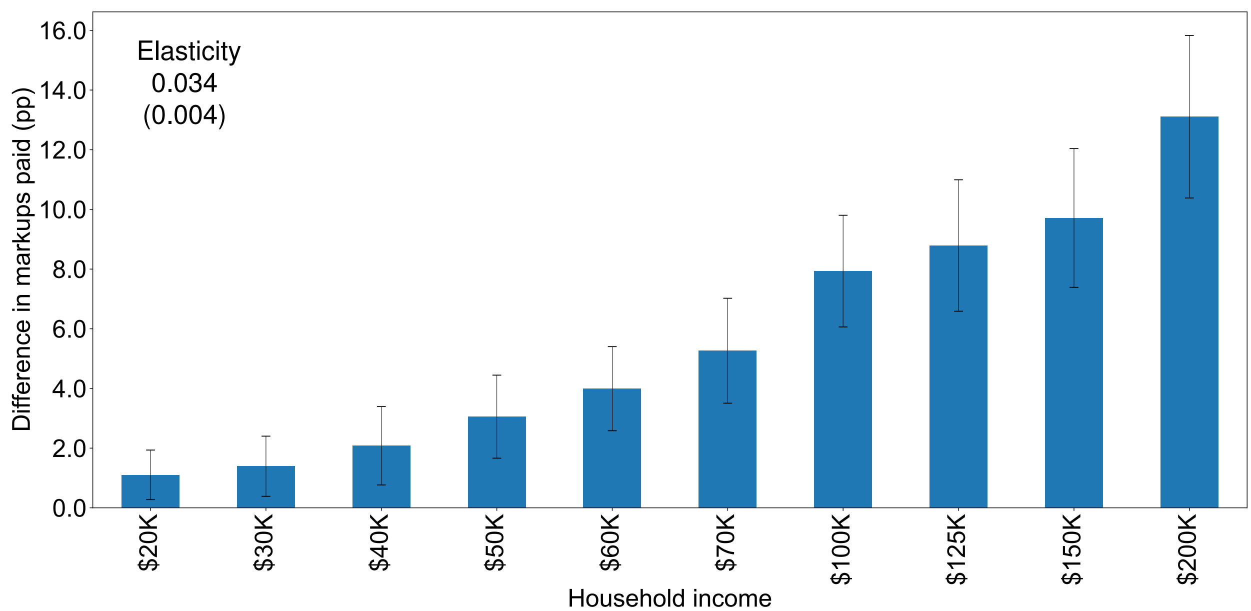
Figure 2a shows that markup gaps estimated by (1) increase systematically with income. In magnitudes, the highest-income households in the sample pay 7 percentage point (pp) higher markups than the lowest-income households after controlling for the

[10] We include these demographic controls to ensure that the measured relationship between markups and income is not driven by the effects of age (studied by Aguiar and Hurst 2007a; Bornstein 2021), race (studied by Butters et al. 2023), or gender (studied by Bhatia et al. 2021) on prices/markups paid.

Figure 2: Relationship between markups paid and household income.



(a) Markup gap within store (sample with store IDs, N = 14.0 million).



(b) Markup gap including cross-store differences, with county and retailer–product controls (N = 25.8 million).

Note: These figures plot the coefficients β_ℓ on household income indicators from specifications (1) and (2). The indicator for the group with income below $20,000 is omitted, so that the plotted coefficients indicate differences relative to households with below $20,000 in income. The elasticity estimates are from regressions (1a) and (2a). All regressions weighted by costs, and standard errors two-way clustered by product brand and household county.

store of purchase.

While (1) controls for unobserved local costs, it omits differences in markups paid across income groups due to differences across stores where households shop. To capture the relationship between markups and income inclusive of differences in markups paid across stores, we estimate the specification,

$$\text{Markup}_{ik} = \sum_{\ell} \beta_\ell 1\{i \text{ has income level } \ell\} + \gamma' X_i + \phi_{c(k)} + \delta' W_k + \epsilon_{ik}. \tag{2}$$

While allowing for comparisons of markups across stores, (2) includes additional controls for unobserved cost shifters: $\phi_{c(k)}$ are county fixed effects that absorb variation in local input costs across space, and W_k is a vector of retailer and product characteristics that control for potential variation in wholesale prices accessed by retailers. Specifically, W_k includes the total sales of the retail chain where transaction k was made (in logs) and interactions between the retailer's sales and total sales of the UPC, the UPC's brand, and the UPC's manufacturing firm (all measured in logs). These controls are intended to absorb potential variation in merchandise costs across retailers, taking into account both the bargaining power of the retailer and of the upstream wholesaler or manufacturer.[11]

Figure 2b shows that, controlling for county fixed effects and the retailer–product characteristics in W_k in (2), the highest-income households in the sample pay 13pp higher markups than the lowest-income households. These are our preferred estimates for the markup differences across income groups, since they account for differences in markups across stores, while controlling for local input costs and potential differences in merchandise costs across retailers.

To summarize the relationships between markups and household income, we estimate the elasticity of markups paid to household income using specifications (1a) and (2a):

$$\log \text{Markup}_{ik} = \beta \log \text{Income}_i + \gamma' X_i + \alpha_{s(k)} + \epsilon_{ik}, \tag{1a}$$

$$\log \text{Markup}_{ik} = \beta \log \text{Income}_i + \gamma' X_i + \delta' W_k + \phi_{c(k)} + \epsilon_{ik}, \tag{2a}$$

where demographic controls X_i, retailer–product characteristics W_k, store fixed effects $\alpha_{s(k)}$, and county fixed effects $\phi_{c(k)}$ are exactly as described in the analogous specifications (1) and (2). For a continuous measure of income, we follow Broda et al. (2009) and recode each

[11] As mentioned above, variation in retailer size should capture most of the variation in merchandise costs across retailers, 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 volume or cost of delivery. The controls in (2) also appear to capture most of the variation in marginal costs across retailers that we recover from demand estimation in Appendix F.2.

household's income as the midpoint of the discrete income bins provided by NielsenIQ.[12] The markup differences in Figure 2 correspond to a micro elasticity of markups to income of 2.2 percent controlling for store of purchase, and 3.4 percent inclusive of cross-store differences in markups paid.

3.1.1 Comparison to previous findings on prices paid for identical products

Seminal papers by Aguiar and Hurst (2007a) and Broda et al. (2009) document that high-income households pay higher prices for identical products. Using NielsenIQ data, Broda et al. (2009) estimate an elasticity of prices paid for identical products (defined by UPC) to income of 1.1–1.4 percent, or 0.9 percent after controlling for store of purchase. To the extent that variation in prices for a UPC is driven by variation in markups, rather than variation in costs over time or across retailers, these differences in prices paid for identical products contribute to differences in markups paid across income groups.

We find differences in markups paid for identical products across income groups, in line with these previous studies, but estimate that this channel accounts for only half of our micro elasticity of markups to household income. When we add store-UPC and county-UPC fixed effects in specifications (1a) and (2a) respectively, Appendix Table B1 reports an elasticity of markups paid for identical products to income of 0.9 percent within store and 1.4 percent overall, similar to the original estimates from Broda et al. (2009).[13,14]

Our estimates for the overall elasticity of markups to income are about two times greater, at 2.2 percent and 3.4 percent respectively. Overall differences in markups across income groups are larger because, in addition to paying higher markups for identical products, high-income households also tend to buy products with higher average markups. The use of cost data to construct markups provides a common unit for cross-product comparisons, allowing us to measure the contribution of basket composition to differences in markups paid for the first time.[15]

[12] For example, a household with reported income in the bin \$12,000–\$15,000 is assigned an income of \$13,500. For the bin with income over \$200,000, we assign an income of \$225,000.

[13] These specifications are nearly identical to Broda et al. (2009), but use log Markup rather than log Price as the dependent variable, since we are interested in decomposing the relationship between markups and income. Using prices instead of markups yields slightly larger estimates of 1.4 percent within store and 1.6 percent overall, because some of the variation in prices is driven by changes in products' costs over the year.

[14] Differences in markups paid for identical products across income groups could be due to differences in exploiting variation in posted prices or differences in coupon usage. In the data, coupons play a negligible role—Appendix Figure B2 shows that coupon savings vary by less than 1pp across income groups—suggesting that differences in markups paid for identical products are predominantly due to differences in the degree to which households exploit spatial and intertemporal variation in posted prices.

[15] In theory, these cross-product comparisons could increase or decrease the markup gap across income groups. For example, classic models of quality discrimination (e.g., Mussa and Rosen 1978; Tirole 1988 Ch. 3) predict that firms set low markups on products bought by high-income customers to deter these

Table 1: Robustness: Micro elasticity of markups to income.

Elasticity of markups paid to income (percent)	Within store	Overall
Baseline	2.2	3.4
Using PromoData base price	2.5	3.6
Using PromoData market-level price	2.3	2.7
Excluding perishable categories	2.0	3.3
With day-of-week fixed effects	2.1	3.4
With supply-side controls	2.1	3.3
Instrumenting for household income	3.8	6.9
With additional retailer–product controls	2.1	3.3

3.1.2 Robustness

We test whether our measures of the micro elasticity of markups to income are robust to alternate measures of retailers' costs, controlling for unobserved spoilage and congestion costs, and to concerns about unobserved volume discounts and selection. In all cases, our estimates of the micro elasticity of markups to income range between 2.0–3.4 percent.

Alternative measures of wholesale costs, spoilage, and congestion costs. Table 1 reports the micro elasticity of markups to income estimated using alternate measures of wholesale costs. Measured elasticities are similar to our baseline results if we use the PromoData base price, which excludes promotional discounts, or if we measure the elasticity of markups to income using only the subset of transactions for which PromoData reports a wholesale price of the UPC in the market of purchase.

We find marginally lower elasticities if we limit the sample to non-perishable items, thus excluding food items that may have higher shipping and spoilage costs, or if we add day-of-week fixed effects to control for changes in congestion faced by retailers over the course of the week. We also test whether the relationship between markups and income is due to products' "supply-side" characteristics by adding controls for the sales shares of the purchased UPC/brand and market concentration in the product category (measured using the Herfindahl-Hirschman Index). Finally, to explore whether markup differences are due to temporary income fluctuations, we instrument for household income with household education and occupation, which are likely to reflect the permanent component of income. Instrumenting for income yields larger estimates, suggesting our elasticities are likely conservative relative to the impact of permanent income on markups paid.

customers from substituting to lower quality products.

Differential volume discounts. While the store fixed effects in specifications (1) and (1a) absorb systematic differences in wholesale, shipping, and local input costs across stores, they do not absorb heterogeneity in costs by product-store pair. This is a problem if some retailers face lower marginal costs on a subset of products, which would be the case for example if large retailers negotiate volume discounts on commodity items but not luxury items. In this case, the data would overstate the marginal cost and understate the markup on commodity items sold at large retailers. If low-income households buy more commodity items at large retailers than high-income households, this mismeasurement would lead us to overestimate the difference in markups paid across income groups.

We address this concern in two ways. First, we test whether the markup gap is driven by large retailers in the sample. We rank retailers by total sales and re-estimate specification (1) 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. Instead, Appendix Figure B1 shows that the estimated markup gap is stable as large retailers are removed from the sample, suggesting that the measured gap in markups paid is not a result of differential volume discounts at large retailers.

Second, we re-estimate specification (1a) including retailer–product controls W_k, which include interactions between retailer size and the total sales of the purchased product, brand, and manufacturing firm, and we further extend W_k to include nonlinear terms (quadratic and cubic) for retailer size. As shown in the final row of Table 1, these additional controls have little impact on our estimated micro elasticity of markups to income.

Selection. Appendix Table A3 shows that unit prices of purchases not matched to the cost data, measured relative to the average unit price of products in the same product category, exhibit a larger covariance with income than those in the merged sample. Since markups tend to increase with unit prices within product categories, this suggests that differences in markups across income groups in our merged sample are likely to be conservative.[16]

[16]In Appendix G.2, we estimate how the markup gap across income groups would change if we observed markups for all Homescan purchases, using unit prices as a proxy for markups and also accounting for the fact that private label products (which we don't have wholesale cost data for) tend to have higher retail markups. Private label reduces the markup gap from 13.1pp to 11.9pp, but accounting for both private label and selection increases the estimated markup gap to 13.4pp.

3.2 Spillovers and the Macro Elasticity of Markups to Income

The *macro elasticity* of markups to income—the relationship between markup and aggregate income—depends on the micro elasticity of markups to income and on spillovers across households. In this section, we develop three specifications to isolate these spillovers, estimating a macro elasticity of markups to income of 8–15 percent.

Micro-to-macro and identification challenges. To fix ideas, suppose the markup paid by household i depend on the household's own (real) income z_i, others' incomes z_{-i}, and a vector of other factors independent of income X. The aggregate markup $\bar{\mu} = \mathbb{E}[\mu_i(z_i, z_{-i}, X)]$, where $\mathbb{E}[\cdot]$ denotes the cost-weighted average across all households. If all incomes rise by a proportion $d \log z$, the change in the aggregate markup to a first order is,

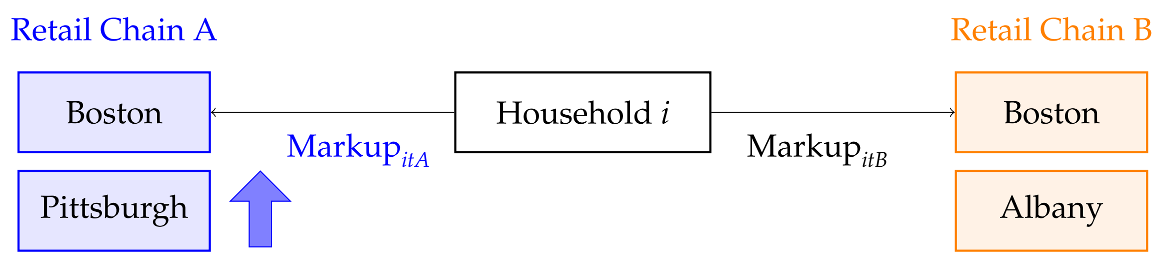
$$\underbrace{\frac{d \log \bar{\mu}}{d \log z}}_{\text{Macro elasticity}} \approx \underbrace{\mathbb{E}\left[\frac{\partial \log \mu_i}{d \log z_i}\right]}_{\text{Micro elasticity}} + \underbrace{\mathbb{E}\left[\frac{\partial \log \mu_i}{d \log z_{-i}}\right]}_{\text{Spillovers}}.$$

The sign of these spillovers is ambiguous ex ante. For example, if markups paid by a household depend on the household's income *relative* to other households, then an increase in others' incomes reduces the household's relative income and its markups paid. In this case, the spillovers would be negative and can even offset the micro elasticity, leading to no aggregate relationship between markups and income. On the other hand, if markups depend only on a household's own income—e.g., if retailers perfectly price discriminate across customers—then these spillovers may be zero, leading the macro and micro elasticities to coincide. Finally, imperfect price discrimination by retailers could lead to positive spillovers and a macro elasticity larger than the micro elasticity.

A first look at the data suggests the last case may be the empirically relevant one. Appendix Figure B3 shows the average markup paid by five income groups split by (a) quintile of county income and (b) quintile of the income of a UPC's other buyers. While the highest-income group pays higher markups than the lowest-income group in all cases, conditional on income, markups paid also increase with the income of other buyers in the same county or purchasing the same product.

There are two primary concerns in interpreting this descriptive evidence. First, these measures of markups may in part reflect unobserved cost shifters, for example due to local input costs. Second, households' choices of where to shop or what to buy may be determined by unobserved households characteristics that are correlated with other buyers' incomes, leading to Manski (1993) reflection problems in estimating spillovers.

Figure 3: Intuition for identifying spillovers by exploiting uniform pricing (eq. (3)).



Identifying spillovers. To overcome these concerns, we develop three specifications to identify spillovers of other buyers' incomes on the markups paid by a household, each of which include store fixed effects (to control for unobserved local cost shifters) and household fixed effects (to control for unobserved household characteristics). These specifications each exploit time series variation in the income of other buyers that a household shops alongside. For this purpose, we compile data on retail markups for all years from 2006 to 2012 using the same procedure described in Section 2.

The first specification exploits the fact that retailers tend to set uniform prices across locations (DellaVigna and Gentzkow 2019) to estimate how the markup paid by a household varies as the income of other buyers at the same retail chain changes over time:

$$\log \text{Markup}_{istk} = \beta_1 \log \text{Income at Retailer Locations}_{r(s),t} + \gamma_{it} + \alpha_s + \phi_{c(s),t} + \varepsilon_{istk}, \quad (3)$$

where Markup_{istk} is the markup paid by household i in transaction k at store s in year t; γ_{it} are household-year fixed effects that control for unobserved household characteristics (and thus the Manski reflection concerns mentioned above); α_s are store fixed effects that control for (static) cost differences across stores; and $\phi_{c(s),t}$ are county-year fixed effects that additionally absorb county-level variation in local costs over time. To calculate average income at a retailer's locations ($\text{Income at Retailer Locations}_{r(s),t}$), for each store s, we take the sales-weighted average of per-capita county income across all stores in the same retail chain $r(s)$ as store s.

Figure 3 provides a stylized illustration of the variation exploited in specification (3). Suppose household i shops at the Boston locations of two retail chains A and B. Retailer A has other locations in Pittsburgh and retailer B has other locations in Albany. If, in a given year, incomes in Pittsburgh rise relative to Albany, how do the markups paid by household i at retailer A's Boston store change compared to at retailer B's Boston store? Since retailers tend to set uniform prices across locations, markups paid may change differentially since the income of buyers household i is effectively pooled with when shopping at retailer A

Table 2: Spillovers of other buyers' incomes on markups paid.

Log Retail Markup	(1)	(2)	(3)
Log Income at Retailer's Locations	0.063**		
	(0.026)		
Log CBSA Income		0.070**	
		(0.012)	
Log Income of Other UPC Buyers			0.142**
			(0.012)
Fixed effects	Household-Year, Store, County-Year	Household-Income, Store, Year	Household-Year, Store-Year
N (millions)	50.9	91.9	97.0
R^2	0.21	0.19	0.21

Note: Regression weighted by sales in 2007 USD, and standard errors two-way clustered by household county and quarter of purchase. ** indicates significance at 5%.

changes relative to the income of buyers household i is pooled with at retailer B.

Table 2 column 1 reports that the elasticity of markups to changes in average income across a retailer's locations, conditional on these controls, is 6.3 percent. That is, for a household shopping at two stores in the same county, doubling the income at locations of one of the store's retail chains is associated with a 6.3 percent rise in the markups paid at that store. This estimate suggests positive and large spillovers of others' incomes on markups paid.

The second specification considers how the markup paid by household i changes as average income in the city where the household lives changes over time:

$$\log \text{Markup}_{istk} = \beta_2 \log \text{CBSA Income}_{m(i,t),t} + \gamma_{i,\text{Income}(i,t)} + \alpha_s + \delta_t + \varepsilon_{istk}. \tag{4}$$

Here, $m(i,t)$ denotes the core-based statistical area (CBSA) where household i lives in year t, and CBSA $\text{Income}_{m(i,t)t}$ is the average income in CBSA m in year t from the Bureau of Economic Analysis (BEA). We again include fixed effects to control for both concerns about unobserved local input costs and unobserved household characteristics: (4) includes store fixed effects α_s; year fixed effects δ_t that absorb secular trends in costs over time; and household–income fixed effects $\gamma_{i,\text{Income}(i,t)}$ that absorb unobserved household characteristics and changes in household income over time. Table 2 column 2 reports the results from specification (4): doubling the per-capita income in a household's city is associated with a 7.0 percent increase in markups paid by the household.

The third and final specification exploits variation in the income of other buyers of a product over time,

$$\log \text{Markup}_{istk} = \beta_3 \log \text{Income of other UPC buyers}_{g(k),t} + \gamma_{it} + \psi_{st} + \varepsilon_{istk}. \tag{5}$$

We calculate the income of other buyers of UPC g in year t as the sales-weighted average income of all households purchasing g in year t excluding household i. This specification admits store-year fixed effects ψ_{st} that allow us to control for time-varying costs at the store level. Table 2 column 3 reports an elasticity of markups paid to the income of a UPC's other buyers of 14.2 percent.

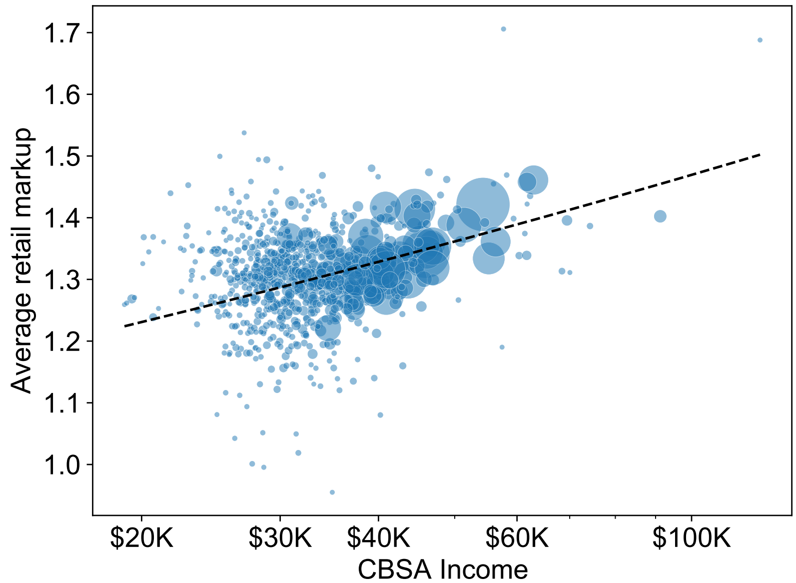
Thus, while all three specifications exploit a different source of variation, Table 2 reports consistent, positive spillovers of others' incomes on markups paid between 6 to 14 percent. The array of specifications allow us to isolate the spillovers of others' incomes on markups while tightly controlling for unobserved household characteristics and local costs. Combining the micro elasticity of markups to own income of 2.0–3.4 percent with these spillovers, we estimate a macro elasticity of markups to income of 8–15 percent.[17]

Comparison to cross-CBSA relationship. While specifications (3)–(5) control for local costs and household characteristics, we can compare the macro elasticity of markups to income estimated from this bottoms-up approach to reduced-form estimates for how markups vary with income across cities. Figure 4 plots retail markups in each CBSA in our data against per-capita income. The elasticity of CBSA markups to income is 11 percent, squarely in the range of macro elasticities we estimate with our bottoms-up approach.

On one hand, this result validates our estimates of the macro elasticity. On the other hand, the fact that the two estimates agree is somewhat surprising, since we might have expected unobserved local costs to bias the elasticity of markups to income measured across cities. In Appendix D, we use data from the Census Retail Trade Survey of Detailed Operating Expenses and data on retail wages and rents to estimate the bias in the elasticity of markups to income across cities due to local costs. This exercise suggests that the bias is likely small, since retailers' labor and rent expenses are less than one-fifth and one-thirtieth of merchandise costs, respectively. We find that if, say, 15 percent of labor and rent expenses were categorized as variable costs rather than overhead costs, the bias in the measured elasticity of markups to income across CBSAs would be about 1pp.

[17] Analogous specifications to (3)–(5) can also be estimated in a single cross-section of data (but without store fixed effects). Estimates from the cross-section are qualitatively similar (see Appendix Table B2).

Figure 4: CBSA per-capita income and aggregate retail markup.



Note: CBSA aggregate markup is the cost-weighted average in our sample, per-capita income is from the Bureau of Economic Analysis (BEA), and bubble size is proportional to CBSA expenditures in our data.

Macro elasticities across space vs. over time. Appendix Table B3 estimates the elasticity of markups to CBSA income using data from 2006–2012, first exploiting variation in income across CBSAs, and second exploiting only variation in income over time within each CBSA. The two estimates are nearly identical, suggesting that the link between markups and income within a CBSA over time is similar to the relationship across CBSAs. Since this exercise is limited to a short sample, we conduct similar analyses using unit prices in each product category as a proxy for markups from 2004–2019. Whether looking across all years or only at long differences between 2004 and 2019 to remove the influence of business cycles, the elasticities across cities and over time are nearly identical. We return to these estimates when discussing balanced growth in the model in Section 4.8.

Heterogeneity across income groups. Appendix Table B6 explores whether the direction and magnitude of spillovers vary across income groups. Such variation could inform our choice of model: for example, in the Varian (1980) model of sales, increasing the share of uninformed buyers (i.e., price-insensitive, high-income households) increases prices paid by uninformed buyers but decreases prices paid by informed buyers (i.e., low-income households). We find that spillovers are positive and of similar magnitude across all income groups, with marginally larger spillovers for high-income households. We return to this evidence when discussing pro-competitive effects in the model (Section 4.9).

4 A Search Model of Income and Markups

In this section, we develop a general equilibrium model that can account for the empirical patterns documented in the previous section. Households in the model differ in two key ways. First, households have different tastes, resulting in differences in basket composition across income groups. Second, households endogenously choose different search intensities. By including both search frictions and differences in tastes across households, the model is able to match the evidence that the micro elasticity of markups to income owes to *both* differences in markups paid for identical products and differences in basket composition. Consistent with our empirical findings, the model also generates spillovers across households and thus different micro and macro elasticities of markups to income.

4.1 Households

A unit measure of households purchases goods indexed by $k = 1, ..., K$. 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 across households. In the calibration, these taste shifters will allow us to flexibly match differences in basket composition across income groups.

When purchasing each unit of consumption, households know the distribution of prices set by firms for each good k, denoted by $F_k(p)$, but do not know which firm sells at which price. Thus, households retrieve price quotes from firms before making a purchase, as in the nonsequential search model of Burdett and Judd (1983). 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 $q_{ik,1}$ household i observes only one price quote, with probability $q_{ik,2}$ the household observes two price quotes, and so on. Each price quote 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 $p \leq R$. The household repeats this search process for each unit of consumption. We assume that units of consumption are infinitesimal, so that integer constraints can be ignored.

We link the distribution $\{q_{ik,n}\}_{n=1}^{\infty}$ to the household's choice of search intensity for good

k, which we denote s_{ik}. Formally, the function $\mathcal{S} : s \mapsto \{q_n\}_{n=1}^{\infty}$ maps search intensity to the distribution of number of price quotes received. Assumptions about the mapping function $\mathcal{S}$ are laid out in Section 4.4, but one should have in mind that greater search intensity increases the likelihood of receiving more price quotes, thereby lowering expected prices.

Households choose these search intensities s_{ik}, as well as time spent working l_i and consumption c_{ik}, to maximize utility subject to a time constraint and budget constraint:

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

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 for good k. With infinitesimal units of consumption, there is no uncertainty in the average price p_{ik} that the household pays for each good k. The budget constraint anticipates that free entry will set firms' profits to zero, so that all earnings come from labor market work.

We assume that the amount of time it takes household i to shop for c units with search intensity s is

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

where a_i is household i's search productivity. Search productivity a_i 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 by adjusting the mapping function $\mathcal{S}$. The relationship between shopping time and consumption amount is more important: if shopping time did not increase with basket size, doubling income would double both the time cost and gains from search, leading to no change in search behavior. However, ample evidence in previous work (Aguiar and Hurst 2007a; Pytka 2018) and in our data (Appendix Tables B4–B5) supports the assumption that shopping time increases with basket size.[18]

The first order condition implies that households increase search intensity s_{ik} as long

[18] Appendix Table B4 shows that search time (measured with proxies from Kaplan and Menzio 2015) increases with basket size, controlling for household income, demographics, and markups paid. Appendix Table B5 shows that within-household increases in basket size over time also increase search time.

as the savings from increasing search intensity are greater than the time cost of doing so:

$$\underbrace{-\frac{\partial p_{ik}(s_{ik}, F_k)}{\partial s_{ik}}}_{\text{Returns to search}} \leq \underbrace{\phi_i}_{\text{Cost of search}}, \tag{7}$$

where the *opportunity cost of search effort* $\phi_i = z_i/a_i$ captures the foregone labor market earnings from increasing search intensity. We focus on the case where each household has an internal solution for s_{ik}, so that (7) holds as an equality.

Aggregate search behavior. We index households by labor productivity $z \in (0, \infty)$, where labor productivity is distributed in the population according to the cumulative density function $H(z)$. We assume that taste shifters and search productivity are identical for all households with a given labor productivity z. Thus, all households with labor productivity z share the same taste shifters $\beta_k(z)$ and 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 z by

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

We 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.

Finally, we use $\Lambda(z)$ to summarize aggregate search behavior for good k, denoted by the probability mass function $\{\bar{q}_{k,n}\}_{n=1}^{\infty}$,

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

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. We take the per-unit variable cost of production as the numeraire (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 $\pi(p) = (p - 1)D_k(p)$, 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 $\{\bar{q}_{k,n}\}_{n=1}^{\infty}$.

Following Burdett and Judd (1983), define a *dispersed-price equilibrium* as an equilibrium in which firm prices follow the distribution $F_k(p)$, all firms choosing a price $p \in \text{supp}(F_k)$

make identical profits, and charging any price $p \notin \text{supp}(F_k)$ results in strictly lower profits. Proving the existence of a dispersed-price equilibrium follows closely from Burdett and Judd (1983); we relegate the details to Appendix C. Given aggregate search behavior $\{\bar{q}_{k,n}\}_{n=1}^{\infty}$ with $\bar{q}_{k,1} \in (0, 1)$, the unique equilibrium price distribution $F_k(p)$ is

$$F_k(p) = \begin{cases} 0 & \text{if } p < \underline{p}_k \\ 1 - \Psi\left[\left(\frac{R-1}{p-1}\right)\bar{q}_{k,1}\right] & \text{if } \underline{p}_k \leq p \leq R \\ 1 & \text{if } p > R, \end{cases} \qquad \text{where} \qquad \underline{p}_k = 1 + \frac{\bar{q}_{k,1}}{\sum_{n=1}^{\infty} n\bar{q}_{k,n}}(R-1), \quad (10)$$

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

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, so that in equilibrium $\pi(p) = f_e$ for all $p \in [\underline{p}_k, R]$. As a result, while changes in the income distribution can affect markups, profits net of entry costs for firms in the model are always constant and equal to zero.

4.3 Equilibrium

An equilibrium is a tuple $\left(F_k, \{c_k(z), s_k(z)\}_{z=0}^{\infty}, M_k\right)_{k=1,\ldots,K}$ such that consumption and search intensity chosen by households for each product, $c_k(z)$ and $s_k(z)$, maximize utility; F_k for each k is a dispersed price equilibrium given aggregate search behavior $\bar{q}_{k,n}$; firms make zero profits net of the entry cost; and 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). We conduct all comparative statics locally around the stable equilibrium in each market.

4.4 Characteristics of the search mapping $\mathcal{S}$

For the analytic results that follow, we make two additional assumptions about the mapping $\mathcal{S} : s \mapsto \{q_n\}_{n=1}^{\infty}$ from search intensity to the distribution of price quotes received.

Assumption 1. Denote the cumulative mass function of $\{q_n\}_{n=1}^{\infty}$ by Q_n. The mapping $\mathcal{S}$ is such that $Q_n(s)$ is weakly decreasing in s for all n and strictly decreasing in s for $n = 1$, and for any non-degenerate distribution F,

$$\sum_{n=1}^{\infty} \frac{d^2 Q_n}{ds^2} \left[\mathbb{E}\left[p|n\right] - \mathbb{E}\left[p|n+1\right]\right] > 0, \tag{11}$$

where $\mathbb{E}[p|n]$ is the expected value of the minimum of n draws from F.

Assumption 1 guarantees that the expected price paid for a good is decreasing and convex in search intensity. First, expected price paid is decreasing in search intensity because increasing s leads to a FOSD shift in $\{q_n\}_{n=1}^{\infty}$. Second, the left-hand side of (11) is the second derivative of expected price paid with respect to search intensity, thereby guaranteeing that expected price paid is convex in search intensity.

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

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

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, the probability of receiving only one price quote $q_1(z)$ is also increasing and convex in z. While Assumption 1 is imposed throughout, Assumption 2 will only be necessary when characterizing how a change in income dispersion affects markups.

While these two conditions may appear technical, we 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; 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. 2023; Menzio 2023).

4.5 Search and markups paid in the cross-section

We first establish how search intensity and markups paid for identical products vary with income in Lemma 1.

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.*

Since $\phi(z) = z/a(z)$, 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 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 prices and markups for identical products.[19]

While the empirical evidence implies that the latter case is relevant for 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 in the auto loan market, where internet access and education likely play a heightened role in consumers' abilities to gather and compare price quotes.[20]

4.6 Micro and macro elasticities of markups to income

While differences in search intensity determine partial equilibrium differences in markups paid for identical products across income groups, the model also features spillovers across households that lead to differences in average markups across products and across economies with different income distributions. To illustrate these spillovers, we consider the case where households are initially identical and explore how a perturbation in incomes affects markups paid.

Proposition 1 (Micro and macro elasticities with identical households)**.** *Suppose Assumption 1 holds, and suppose households are initially identical, so that $z_i = z$, $a_i = a$, and $\beta_{ik} = \beta_k$ for all households. In response to a perturbation to household i's labor productivity, dz_i, and to the labor productivity of all other households dz_{-i}, the change in markups paid by household i is*

$$d\mu_i = \left(1 - \frac{d\log a}{d\log z}\right)\left(\underbrace{\kappa_1 dz_i}_{\substack{\textit{Own income:}\\ \textit{Through search choice}}} + \underbrace{\kappa_2 dz_{-i}}_{\substack{\textit{Spillovers:}\\ \textit{Through price distribution}}} + \underbrace{\kappa_3 \frac{\partial s_i}{\partial s_{-i}} dz_{-i}}_{\substack{\textit{Spillovers:}\\ \textit{Through search choice}}} \right),$$

where $\kappa_1, \kappa_2, \kappa_3 > 0$ are strictly positive constants defined in Appendix C.4.

[19] In the data, households pay different prices for identical products due to exploiting both spatial and intertemporal variation in prices. Since the model is static, price dispersion in the model is entirely spatial (across stores). We interpret spatial dispersion in the model as a stylized representation of both types of variation: since firms are indifferent between prices in supp(F), they may randomize over prices in $F(p)$ from week to week (as they do in the Varian 1980 model of sales). Households can retrieve multiple quotes both by visiting the same store in different weeks and by visiting different stores.

[20] In 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.

First, note that how markups paid by household i respond to changes in its own income depend on whether search productivity $a(z)$ rises more or less than one-for-one with labor productivity z. As discussed in Lemma 1, if $a(z)$ rises less than one-for-one with z, then an increase in household i's income z_i increases the markups paid by household i.

Markups paid by household i also respond to changes in other households' incomes, through two channels. First, if $a(z)$ rises less than one-for-one with z, then an increase in others' incomes leads to a decrease in average search intensity. This decrease in search intensity has a direct effect on markups paid by household i because it affects the prices set by firms: when search intensity falls, firms set higher markups, increasing the markup paid by household i. Second, the decrease in average search intensity also affects the *returns* to search, which in turn impacts the optimal level of search intensity for household i. We show in Appendix Lemma 5 that when search costs are sufficiently low, household search decisions are strategic substitutes ($\partial s_i/\partial s_{-i} \leq 0$). This means that a decrease in aggregate search intensity leads household i to increase its own search intensity, partially offsetting the increase in markups paid by household i. These strategic interactions create a moderating feedback loop between aggregate income and search intensity, dampening the macro elasticity of markups to income. This mechanism also offers an explanation for findings from Nevo and Wong (2019) that the Great Recession led to both increases in household search intensity and a decline in the returns to search effort. Together, the spillovers of others' incomes on both prices and search choices lead the macro elasticity of markups to income in the model to differ from the micro elasticity across households.

4.7 Comparative statics of the aggregate markup

We now proceed to characterize how changes in the income distribution affect the aggregate markup in the economy. First, we prove an intermediate result that links aggregate markups in the model to the fraction of customers receiving only one price quote.

Lemma 2. *Define the aggregate markup for product k, $\bar{\mu}_k$, as the ratio of total sales to total variable costs for good k, and define the aggregate markup $\bar{\mu}$ as the ratio of total sales to total variable costs in the economy. In equilibrium, $\bar{\mu}_k = 1 + (R-1)\bar{q}_{1,k}$, and $\bar{\mu} = 1 + (R-1)(\sum_k C_k \bar{q}_{k,1} / \sum_k C_k)$.*

This result is useful because it allows us to characterize changes in the aggregate markup without keeping track of the entire distribution of customers and their search intensities. The intuition is that, 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 with price R are those that receive no other price quotes, the fraction of customers receiving only one price quote $\bar{q}_{k,1}$

is a sufficient statistic for the profits of firms selling good k. Thus, the fraction of customers receiving one quote $\bar{q}_{k,1}$ allows us to calculate the aggregate markup for each good k.

We use this result to characterize how changes to the income distribution affect the aggregate markup. For these results, we specialize to the case where $K = 1$, so that there is a single distribution of buyers' incomes $\Lambda(z)$.

Proposition 2 (First-Order Shift)**.** *Suppose $K = 1$ and Assumption 1 holds. Consider a perturbation in the household distribution such that the new $\tilde{\Lambda}(z)$ first-order stochastically dominates the initial $\Lambda(z)$. This perturbation increases the aggregate markup if $\phi(z)$ is **increasing** in z.*

Proposition 3 (Mean-Preserving Spread)**.** *Suppose $K = 1$ and Assumptions 1 and 2 hold. Consider a perturbation in the household distribution such that the new $\tilde{\Lambda}(z)$ is a mean-preserving spread of the initial $\Lambda(z)$. This perturbation increases the aggregate markup if $\phi(z)$ is **increasing and convex** in z.*

In both Propositions 2 and 3, the effect of a change in the income distribution on markups hinges on how the opportunity cost of search effort $\phi(z)$ varies with income. The two conditions on $\phi(z)$ determine whether the first-order stochastic shift in Proposition 2 or the mean-preserving spread in Proposition 3 raises customers' average cost of search effort. An increase in the average cost of search effort in turn increases the share of customers that receive only one price quote, and thus, by Lemma 2, leads to a rise in the aggregate markup.

4.8 The race between search and labor productivity

While Proposition 2 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, 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. In other words, whether growth leads to an increase in the aggregate markup in the model depends on a *race between labor and search productivity*. If labor productivity growth outpaces search productivity growth, opportunity costs of search effort rise over time, leading to rising markups (tending in the limit toward the Diamond 1971 monopoly price equilibrium).

This result relates to a literature that asks why price dispersion and markups have not fallen over time, despite apparent improvements in search technology. Menzio (2023) reviews an empirical literature showing constant price dispersion over time and proposes a model in which declining search frictions coincide with constant price dispersion due to increased specialization by sellers. This paper makes the complementary point that increases in search productivity need not lead to a decline in price dispersion or markups—and can in fact coincide with increasing markups—if labor productivity is also rising.

4.9 Discussion

Taking stock. The model developed in this section can account for each of the empirical patterns we document in Section 3: (1) households across income groups pay different markups; (2) differences in markups across income groups owe to both differences in exploiting variation in prices for identical products and differences in basket composition; and (3) spillovers across households lead the macro elasticity of markups to income to differ from the micro elasticity. As we will show in the next section, the model can in fact match the quantitative magnitudes of each of these patterns in the data.

It is useful to highlight how this model differs from models that instead attribute differences in price sensitivity to non-homothetic preferences or differences in utility primitives across households (e.g., Berry et al. 1995; Simonovska 2015; Handbury 2021; Auer et al. 2022). While those models can be calibrated to match differences in markups across income groups, they are unable to account for the fact that a substantial part of the markup gap across income groups in our data is due to differences in the extent to which households exploit variation in prices for identical products. Moreover, since price sensitivity in those models are determined by utility primitives rather than endogenous search decisions, they lack the moderating feedback loop between aggregate income and price sensitivity in the search model. In a previous version of this paper, we calibrated a model with non-homotheticities in the elasticity of substitution and taste for quality to match differences in markups paid across income groups. Due to the absence of this moderating feedback loop, that model substantially overshot the macro elasticity of markups to income that we estimate in the data.

Extension with 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. One explanation for their findings is that an increase in market size leads to a reduction in relative prices due to pro-competitive effects of firm entry on markups.

We can incorporate these pro-competitive effects by adding an endogenous response of search productivity to entry:

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

Intuitively, as more firms enter to supply good k, households can collect more price quotes for k in the same amount of search time. The elasticity of search productivity to the mass of firms, ζ, parameterizes the strength of this effect. Appendix Figure B11 shows that $\zeta = 0.3$ can match the elasticity of prices to market size from Jaravel (2019).

Our data on retail markups do not support a choice of $\zeta > 0$, however. Appendix Table B6 shows that, both in the cross-section of cities and 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 $\zeta > 0$ would instead imply that markups for high-income households rise more slowly with aggregate income, due to offsetting pro-competitive effects for goods purchased by high-income households.[21] Thus, our baseline results assume $\zeta = 0$.

5 Calibration

In this section, we calibrate the model to match differences in basket composition and markups paid across income groups, and explore how the model performs in matching untargeted moments from the data and the literature.

5.1 Calibration procedure

Households in the model differ in two key ways: households have different tastes, generating differences in basket composition across income groups, and have different opportunity costs of search effort. We choose household taste shifters to exactly match households' expenditures across goods, and we choose opportunity costs of search effort to exactly match the differences in markups paid across income groups.

[21] The difference in results may be due to our focus on retail markups rather than price indices (compared to Handbury 2021) and our use of difference-in-differences across cities to discipline these spillovers (compared to the aggregate time series evidence in Jaravel 2019).

Table 3: Calibration parameter values and sources.

Parameter		Value	Description
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	$3.0^{\dagger}$	98th percentile of markups in the data
Search mapping	$\mathcal{S}$	Poisson	Albrecht et al. (2023), Menzio (2023)
Opp. costs of search	$\phi(z)$	-	Match markup differences from Figure 2b
Search productivity	$a(z)$	-	Solved from $\phi(z) = z/a(z)$
Household distribution	$H(z)$	-	Saez and Zucman (2019)

Note: † indicates that Online Appendix Table B7 reports robustness to alternate values.

Preferences and taste shifters. We first choose preferences and taste shifters across income groups to match differences in expenditure shares across different products (UPCs) in the data.[22] To simplify computation, we aggregate UPCs in the data into $K = 10$ groups and define each group as a good in the model. We order UPCs in the data by average buyer income, to capture differences in the composition of their customers, and split them into 10 groups with equal sales. Appendix Figure B4 shows the fraction of expenditures of each income group on each of these 10 product groups. We then choose parameters to exactly match these spending shares: we assume preferences over goods are Cobb-Douglas ($\sigma = 1$) and choose taste shifters, $\beta_k(z)$, to match the share of expenditures by households with income z on good k.

Note that the choice of Cobb-Douglas preferences is without loss of generality in the cross-section, since for any other value of σ, we could adjust the taste shifters $\beta_k(z)$ to continue matching observed expenditure shares. However, the assumption of Cobb-Douglas preferences can matter for counterfactuals, since these preferences will determine how spending shares evolve as relative prices across goods change. We show in Appendix Table B7 that choosing other values for σ has little effect on our results.[23]

[22]Calibrating basket composition using household expenditure shares across UPCs assumes that the same UPC sold across different stores is an identical good from the perspective of households. This assumption is in line with previous work on differences in prices paid for identical products (e.g., Aguiar and Hurst 2007a), but we readily acknowledge that one could draw the line more or less expansively, e.g., defining a unique product as a retailer–UPC pair or bundling together UPCs that contain the same product in different package sizes. Ultimately, since basket composition determines the extent of market segmentation and thus the spillovers across households, we judge this calibration choice by whether the model generates spillovers in line with our empirical estimates (see Table 4).

[23]We also do not find changes in segmentation from 2004–2019 that would inform a different choice of σ. Boar and Giannone (2023) also find little evidence of changes in consumption segregation over time.

Search parameters. Next, we choose opportunity costs of search effort $\phi(z)$ to exactly match differences in markups paid across income groups from Figure 2b. Following Albrecht et al. (2023) and Menzio (2023), we assume the mapping function $\mathcal{S} : s_i \mapsto \{q_{i,n}\}_{n=1}^{\infty}$ is Poisson. We take firms' unit costs as the numeraire and set households' reservation price to the 98th percentile of markups in the data, $R = 3.0$. We then calibrate opportunity costs of search effort $\phi(z)$ using the following two-step procedure:

1. **Inner loop: Price distributions F_k.** Given a guess for $\phi(z)$, the price distribution F_k and household search decisions $s_k(z)$ solve a fixed point for each k. The fixed point comes from the fact that, for a given $s_k(z)$, F_k is pinned down by (10), and given F_k and $\phi(z)$, households' first order conditions (7) pin down search intensities $s_k(z)$. We iterate this loop until the price distributions F_k and search intensities $s_k(z)$ converge.

2. **Outer loop: Opportunity costs of search effort $\phi(z)$.** Given an initial guess $\phi(z)$, we calculate price distributions F_k for all k. Given F_k and households' expenditure shares, there is a monotonic mapping from households' search costs to markups paid. We re-compute $\phi(z)$ to match an aggregate markup of 1.32 and the gaps in markups paid across income groups from Figure 2b. We iterate this process until $\phi(z)$ converges.

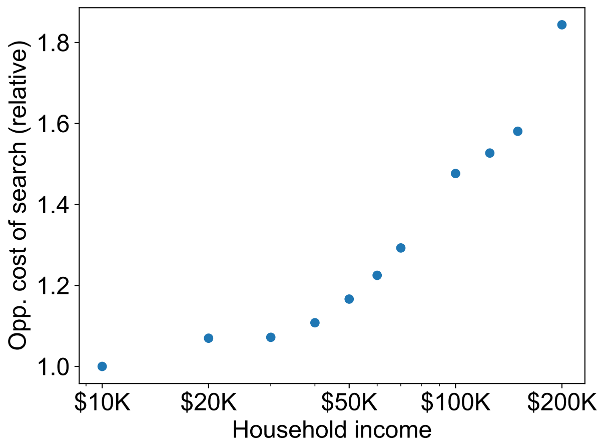
Finally, for the distribution of households across income groups, we use estimates of pre- and post-tax income by percentile of the income distribution from 1950 to 2018 from Saez and Zucman (2019). We assume markups paid and search behavior of any household with income over \$200K are equal to the \$200K income group, so that the results are not influenced by extrapolation beyond the income range observed in the data. Our baseline results assume a unitary elasticity of expenditures to post-tax income; however, using the expenditure shares of income groups on NielsenIQ or food-at-home categories produces similar quantitative results (see Appendix Table B7).

5.2 Calibrated statistics

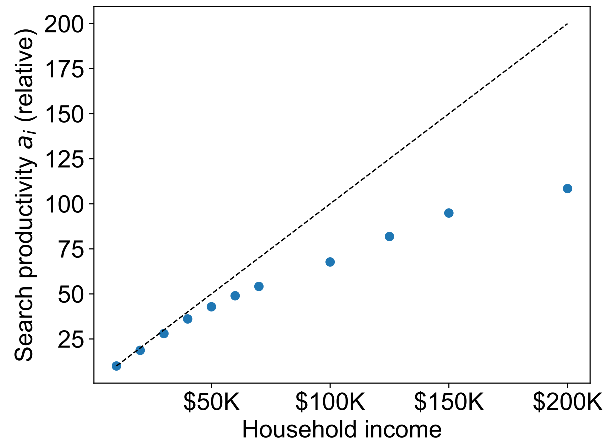
Figure 5 plots calibrated values for opportunity cost of search effort and search productivity by income group. The left panel shows that the calibrated opportunity costs of search effort $\phi(z)$ are increasing and convex in log income, satisfying the conditions in Propositions 2 and 3. This is because search productivity $a(z)$, shown in the right panel, increases less than one-for-one with income, particularly at higher income levels.

The calibrated model matches several untargeted moments from the data and existing literature, including the sources of the markup gap across income groups, the relationship

Figure 5: Opportunity cost of search effort and search productivity by income.



(a) Opportunity cost of search effort $\phi(z)$.



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

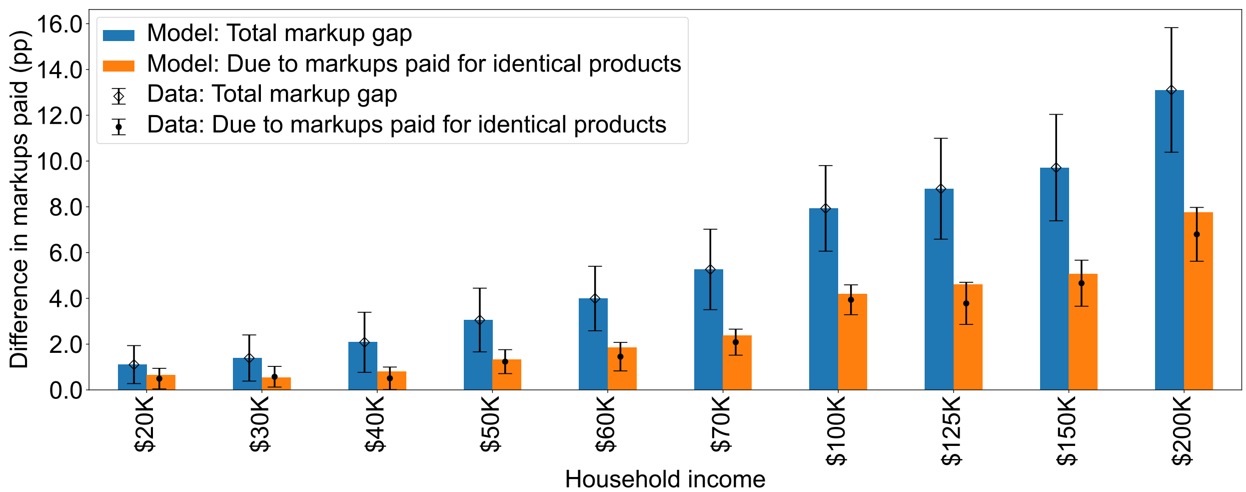
between household income and search intensity, and spillovers of other buyers' incomes on markups paid and search intensity.

Sources of the "micro elasticity" of markups to income. As discussed in Section 3.1, differences in markups paid across income groups owe to both differences in markups paid for identical products and to differences in basket composition. Figure 6 plots markup differences across income groups and the portion due to differences in markups paid for identical products. In both the model and data, differences in markups paid for identical products are responsible for about half of the overall markup gaps across income groups, with the remainder due to differences in basket composition, suggesting that the model captures how both channels contribute to markup differences by income in the data.

It is worth emphasizing that in the model, the *reason* that products consumed by high-income households have higher average markups—and thus that basket composition increases the markup gap across income groups—is because those products have high-income customer bases. This contrasts with previous models that explain higher markups on goods consumed by high-income households due to essentially intrinsic features of the products, such as quality (e.g., Fajgelbaum et al. 2011). Across the ten products in our calibration, the elasticity of products' average markups to the average income of their buyers is 10 percent.[24] Quantitatively, these differences in average markups across products, combined with differences in households' expenditure shares, match the contribution of basket composition to the markup gap across households in the data.

[24] In the data, the elasticity of the average markup for a UPC to average buyer income is 28 percent overall and 10 percent after controlling for the product group.

Figure 6: Sources of the markup gap across income groups: Model vs. data.



Note: The bars are markup differences across income groups in the model, and the scatterpoints are estimates from the data (from (2) and a variant of (2) that adds UPC fixed effects).

Spillovers and the "macro elasticity." Next, we check whether the model is able to match spillovers across households and the macro elasticity of markups to income. These moments are especially important for our counterfactuals, since the macro elasticity of markups to income determines how changes in the income distribution affect aggregate markups predicted by the model. We simulate the model with the income distributions of 881 CBSAs (described in Section 6) and run regressions of markups paid and search intensity on own income and others' incomes analogous to our regressions in the data.

For markups, the model generates a "micro elasticity" of markups to income of 2.8 percent, in line with our estimates of 2.0–3.4 percent in Section 3.1. The elasticity of markups paid to others' incomes, 6.1 percent, is at the low end of our spillover estimates from Section 3.2. Thus, the macro elasticity of markups to income in the model is about 9 percent, squarely in the 8–15 percent range estimated in the data.

Search intensity in the model also matches patterns in the data. The elasticity of search intensity, measured as the average number of price quotes received per purchase, to income is −0.11. Using various measures of search intensity from Kaplan and Menzio (2015), Appendix Table E1 finds elasticities of search intensity to income in the data between −0.08 and −0.40. Differences in search intensity in the model are also consistent with previous work by McKenzie and Schargrodsky (2005), who find that, controlling for quantity purchased, shopping frequency for high-income households is 30 percent lower than low-income households: their estimates correspond to an elasticity of about −0.12. In the model, households also exert more search effort when others' incomes rise,

Table 4: Elasticity of markups paid and search intensity to own and others' incomes.

	Log markup		*Search intensity*	
	Data	Model	Data	Model
Log Own Income	0.038	0.028	−0.26	−0.11
Log Others' Income	0.083	0.061	0.03	0.03

Note: The first column is from Appendix Table B2, and the third column is from Appendix Table E1. Model columns are from simulations of income distributions for 881 CBSAs.

consistent with our estimates in the data. Finally, returns to search in the model—doubling a household's search intensity decreases prices paid by 7–9 percent—line up closely with Aguiar and Hurst's (2007a) estimates on the returns to shopping effort.

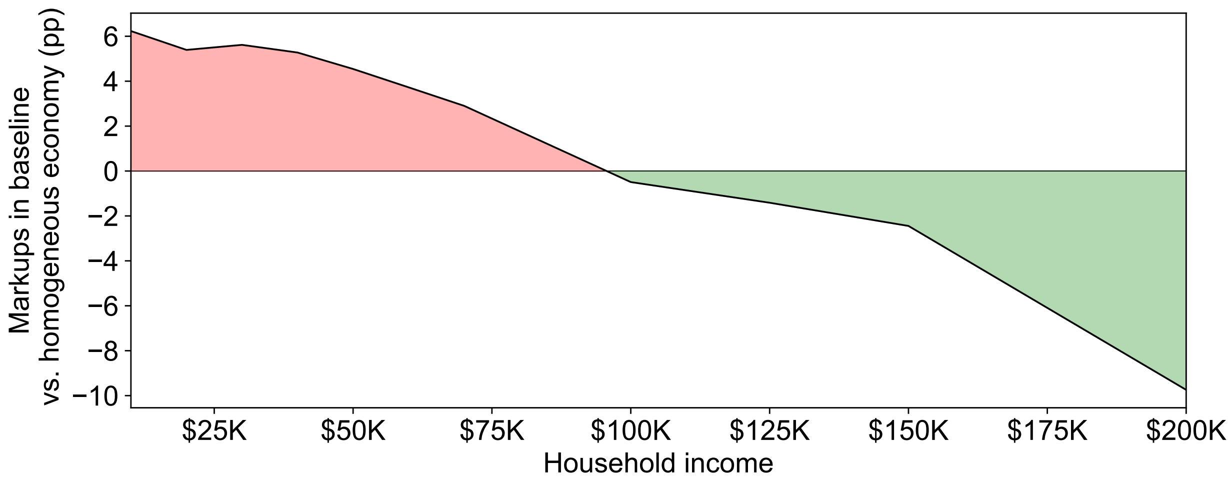
6 Implications

We first use the calibrated model to explore spillovers across households and the distributional impact of income inequality across households. We also simulate the model with the income distribution of different U.S. cities, finding that the model has much more success in replicating the behavior of retail markups across cities than standard supply-side models that predict markups using data on market shares and concentration.

Spillovers across income groups. In the model, households' search decisions affect prices set by firms and hence the markups paid by other households. To quantify these spillovers, for each income level z, we consider an economy in which there is a representative household of that income. We compare the aggregate markup that each household would pay in that economy to the markup it pays in the baseline calibration. Besides measuring the extent to which households benefit from or are harmed by the presence of households with incomes other than their own, these estimates also describe the distributional impact of limits on retailers' abilities to price discriminate across income groups.

Figure 7 shows that low-income households would pay 6pp lower markups in an economy populated with only low-income households, while the highest-income households would pay 10pp higher markups in an economy with only high-income households. These differences translate into meaningful savings or losses. For example, for low-income households, a 6pp reduction in markups corresponds to savings of over $250 on the $5,000 that low-income households spend annually in the data.

Figure 7: Markups paid relative to homogeneous income economies.

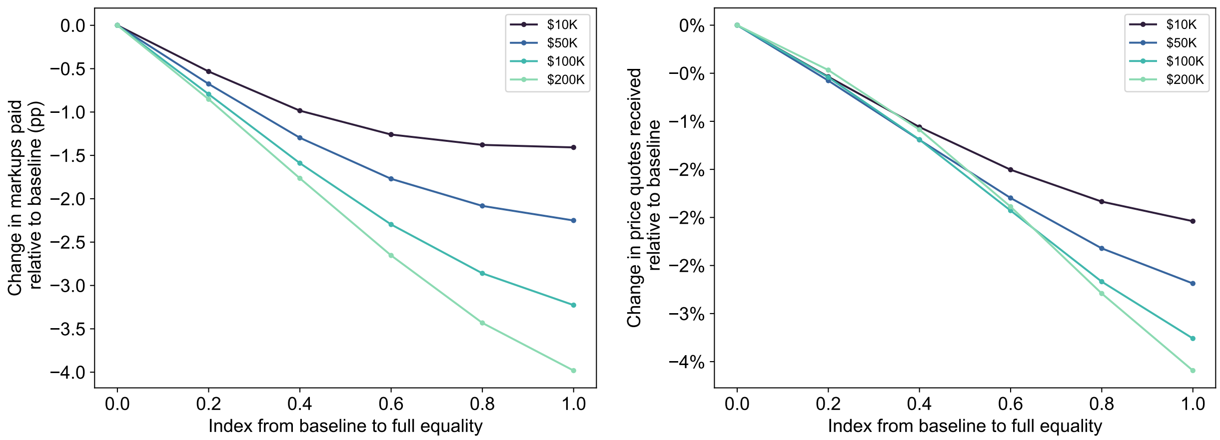


Besides affecting markups paid, the composition of households in the economy also affects search decisions since, when incomes rise, households compensate by exerting more search effort. Appendix Figure B5 shows that low-income households retrieve 5 percent more quotes per purchase in the baseline economy than they would in an economy with other households of their own income level. Together, differences in markups paid and search time shape the welfare impact of imperfect price discrimination by retailers across households: Appendix Figure B5 shows that, in the baseline equilibrium compared to in an economy with perfect price discrimination, low-income households are 5 percent worse off, while high-income households are 6 percent better off.

Effects of income dispersion. To explore the effects of inequality on households, we consider how markups paid and search time would change if income inequality were reduced. We simulate the model under counterfactual household distributions that scale all household incomes toward the average per-capita income in the economy.[25] Figure 8 plots markups paid and search intensity for four income groups as we change income dispersion from its baseline level to full equality. Reducing income dispersion leads to a decline in markups and search intensity for all income groups. The intuition for this result goes back to the convexity of opportunity cost of search time: 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.

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

Figure 8: Effects of changing income inequality on markups and search behavior.



(a) Markups paid. **(b)** Avg. price quotes received.

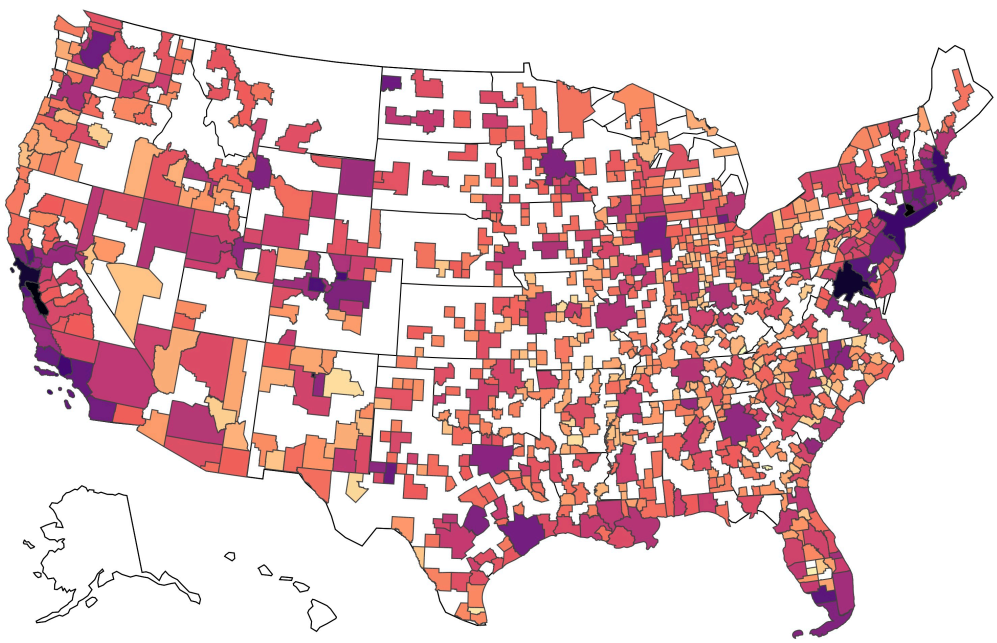
Interestingly, the effects of reducing income dispersion on markups paid and search intensity are most pronounced for high-income households. In Appendix Figure B6, we show that the differential effects of income dispersion on high-income households are due to market segmentation: these differences largely disappear in a single-product calibration of the model. Income dispersion leads to larger increases in markups paid by high-income households because low search-intensity households that enter the economy buy goods that are predominantly purchased by high-income households.

Markups across cities. Finally, we explore how differences in income distribution shape markups across U.S. cities, and compare the predictions of our model to both retail markups in the data and predictions from a "supply-side" model of markups. Our data on the income distribution in each CBSA comes from the American Community Survey five-year estimates. We simulate the model pairing each CBSA's income distribution with the search and taste parameters, $\phi(z)$ and $\beta_k(z)$, calibrated for the national data.

Figure 9 shows the predicted CBSA markups. The model predicts high markups in high-income coastal cities, such as New York, Washington, D.C., Boston, and San Francisco. Some less dense, high-income areas are also predicted to have high markups, such as Bridgeport, Connecticut; Napa, California; and Jackson, Wyoming. These patterns mirror the pattern of markups across CBSAs in the data (Appendix Figure B7).

The first two columns of Table 5 show that per-capita income and income inequality—as measured by Gini Indices from the ACS—are important determinants of the CBSA markups predicted by the search model. These predictions are borne out in the data: the

Figure 9: 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 45 percent (dark purple).

aggregate (cost-weighted) markup across CBSAs in our data rises with both per-capita income and inequality. Moreover, column 5 shows that markups predicted by the model account for 31 percent of the variation in CBSA markups in the data. For comparison, a linear regression of CBSA markups on per-capita income and inequality explains only 28 percent of the variation in the data, suggesting that the model successfully summarizes other moments of the income distribution that matter for markups in the data.

We also calculate CBSA markups predicted by a "supply-side" nested CES model. The nested CES model is frequently used to infer markups from data on firm market shares and concentration (e.g., Smith and Ocampo 2023). Predictions from our search model isolate how demand composition affects markups, while predictions from the nested CES model focus on how supply-side factors—market shares and concentration—affect markups.[26] We find that markups predicted by the nested CES model account for only 10 percent of the variation in CBSA markups and in fact covary negatively with markups

[26]To calibrate the nested CES model, we 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 goods. We choose $\rho = 5$ and $\eta = 2$ to match the level and dispersion of markups in the data. We have found similar results experimenting with different values of ρ and η and with Cournot vs. Bertrand competition.

Table 5: Markups across cities in the model and data.

	Model-Predicted		Data				
Log CBSA Markup	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log CBSA Income	0.086**	0.081**	0.110**	0.102**			0.019
	(0.001)	(0.001)	(0.006)	(0.007)			(0.015)
Gini Index		0.088**		0.153**			0.075
		(0.011)		(0.057)			(0.058)
Log Model-Predicted Markup					1.248**		0.956**
					(0.063)		(0.169)
Log Nested CES Markup						-0.720**	-0.123*
						(0.072)	(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 outcome variable in columns 1–2 is the aggregate markup predicted by our search model given CBSA income distributions, and in columns 3–7 is the aggregate CBSA markup in the data. Regressions weighted by total CBSA expenditures. ** is significant at 5%, * at 10%.

in the data. As shown in column 7, adding these "supply-side" model markups and income measures does not meaningfully increase the share of variation in CBSA markups explained compared to the search model-predicted markups alone.

7 Changes in the Income Distribution from 1950–2018

Our main counterfactual considers how changes in the income distribution from 1950 to 2018 contribute to changes in the aggregate retail markup, holding all other factors constant. We first use moments from the data to provide model-free estimates, then use our calibrated model. Through the lens of the model, changes in the income distribution from 1950 to 2018 account for a 11pp rise in the aggregate retail markup.

7.1 Model-free estimates

In Section 3.2, we estimate a macro elasticity of markups to income of 8–15 percent. Given an aggregate markup of 32 percent, since post-tax real income grew 3.5 times from 1950 to 2018, a back-of-the-envelope calculation suggests a change in the aggregate markup of $1.32 \times \log(3.5) \times (0.08 \text{ to } 0.15) \approx 13.2 \text{ to } 23.2$pp.

This estimate does not account for non-linearities in the aggregate relationship—which are likely to be matter, since the aggregate markup must be bounded below by one—or for changes in income dispersion. Hence, we now turn to the model.

7.2 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. 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. So, our baseline assumption for this counterfactual 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. Thus, for income distributions from 1950 to 2018, we assume that households with real post-tax earnings z have the same search productivity $a(z)$ and opportunity cost of search effort $\phi(z)$ as households with those real post-tax earnings in 2007.[27]

The solid line in the left panel of Figure 10 plots the predicted aggregate retail markup using the distribution of post-tax, real income from Saez and Zucman (2019) from 1950 to 2018. Over this period, the model predicts a 11pp 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 dashed line in the left panel of Figure 10 plots the predicted retail markup holding income dispersion fixed 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 2.5pp lower than at the 2018 level of income dispersion. Table 6 summarizes these predictions.

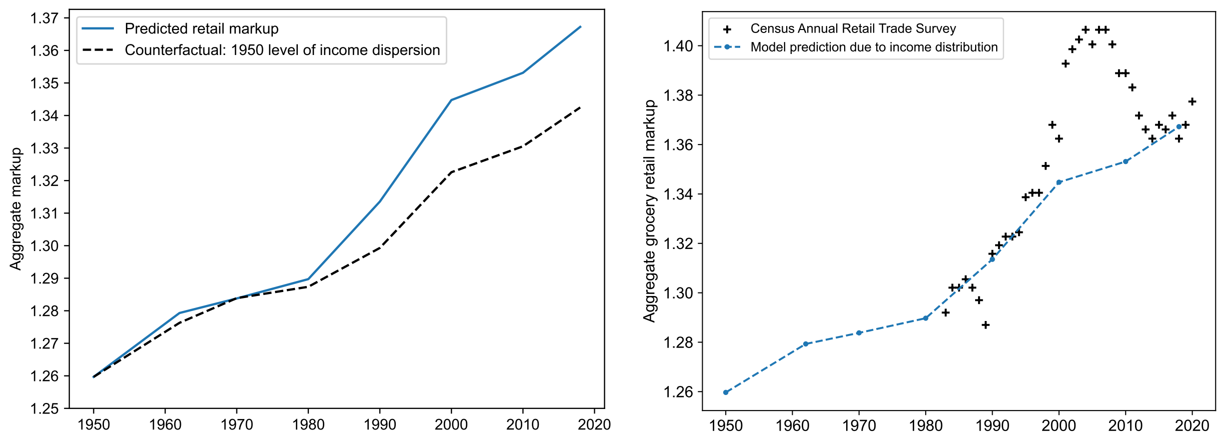
How does the rise in markups compare to data? The right panel of Figure 10 compares the path of the aggregate markup in the model to data on gross margins for grocery stores from the Census of Annual Retail Trade Survey. An aggregate markup calculated from grocery store gross margins reported by the Census increases from 29 percent in 1983 to 38 percent in 2020.[28] The path of the aggregate retail markup predicted by the model (the dashed blue line) tracks closely with the secular trend in markups in the data.

Notably, markups in the data exhibit a rise and fall in the late 2000s. While changes in the income distribution explain the secular trend in markups well, they are not sufficient to explain this boom and bust in markups. Stroebel and Vavra (2019) document a link

[27] Since $a(z)$ is increasing in z, search productivity improvements are "baked in" to this exercise. Appendix Figure B8 shows that long-term trends in search time in the model broadly resemble time use surveys from Aguiar and Hurst (2007b), and that price dispersion in the model is roughly constant (slightly declining) over time (offering an explanation for constant price dispersion over time in the data—see Menzio 2023).

[28] Appendix Figure B13 shows that other sectors in the Census ARTS also exhibit an upward trend in gross margins. Using Compustat data on public retail firms, Anderson et al. (2018) find that gross margins rose from 0.27 in 1982 to 0.31 to 2014. Assuming constant returns, their estimates correspond to a similar 8pp rise in markups from 37 to 45 percent.

Figure 10: Predicted aggregate retail markup with 1950–2018 income distributions.



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 from the Census Annual Retail Trade Survey, calculated from gross margins under the assumption of constant returns.

between retail markups and housing wealth; we find that applying their estimates of the elasticity of markups to housing wealth bridges the gap between our model's predictions and the data for this period (see Appendix Figure B10). These results suggest that other mechanisms such as housing wealth may be important for markups over the business cycle, but that long-run changes in the income distribution can account for the long-run rise in retail markups in the data.

The role of reallocations. Changes in the aggregate markup reflect both changes in the markups set by individual firms and compositional shifts across firms. Autor et al. (2020), Baqaee and Farhi (2020), and De Loecker et al. (2020) suggest that reallocations to high-markup firms have played a central role in increasing markups in the U.S. economy.

Table 6 reports that, in our model, 56 percent of the rise in the aggregate retail markup over time is due to within-firm changes in markups, while 44 percent is due to reallocations of sales across firms and products. These reallocations occur for two reasons: growing incomes lead households to shift their baskets to high-markup products, and within products, high-income households search less and thus buy more often from firms with relatively high markups. Appendix Figure B9 shows that these changes lead to declining sales for firms at the bottom of the markup distribution for a product and growing sales for firms at the top of the markup distribution. These reallocations emerge in the model as an endogenous outcome of changes to the demand side of the economy.

Table 6: Predicted change in aggregate retail markup from 1950–2018.

Period	Predicted Δ in markup	Due to Δ Income level	Due to Δ Income dispersion	Due to Changes in markups	Due to Reallocations of sales
1950–2018	10.8pp	8.3pp	2.5pp	6.1pp	4.7pp
1950–1980	3.0pp	2.8pp	0.2pp	1.9pp	1.1pp
1980–2018	7.8pp	5.5pp	2.2pp	4.2pp	3.5pp

Implications for efficiency. There are three sources of inefficiency in the decentralized equilibrium: (1) households face heterogeneous markups across goods, distorting their consumption choices; (2) positive markups lead to excessive firm entry, since firms selling the same good are perfect substitutes to households; and (3) price dispersion leads households to waste time on search rather than labor market work. Appendix Figure B12 computes welfare gains of 30–40 percent from moving to the efficient frontier. The distance to the frontier in the model is also growing over time, primarily due to rising markups and thus more excessive entry. However, it is worth stressing that the rise in markups in the model is in fact a consequence of labor productivity and the efficient level of output rising over time. In other words, rising markups reflect household incomes and welfare growing over time, but also suggest larger potential gains for moving to efficiency.

Implications for consumption inequality. As stressed by Aguiar and Hurst (2007a), differences in expenditures may be a poor proxy for differences in consumption if households pay different prices for identical products. Appendix Table B8 reports that the Gini index of costs of goods purchased—a proxy for consumption inequality—is 2 percent lower and has grown more slowly than the Gini index for post-tax income in the model. These estimates should be interpreted with caution, since they assume that differences in markups across income groups in our sample extend to the entire consumption bundle.

Robustness to calibration choices. In Appendix Table B7, we report how our results change when we vary the number of goods K, allow for pro-competitive effects of firm entry, vary the household expenditure shares we use for calibration, and relax the assumption of Cobb-Douglas preferences across goods. We find broadly similar results when varying each of these parameters: in all cases, we find that the model predicts a rise in markups between 9.5–14.3pp from 1950 to 2018.

8 Discussion and Broader Questions

Both our model-free and model-based estimates suggest that changes in the income distribution over time can account for a substantial part of the rise in retail markups in the data, without requiring additional changes to the supply side of the economy. The fact that additional supply side forces may not be needed to account for the observed rise in markups appears consistent with our finding that market shares and concentration have limited explanatory power for patterns in retail markups across cities, and with recent evidence that local market concentration in retail may actually be falling over time, rather than rising (Benkard et al. 2021; Rossi-Hansberg et al. 2021). At the same time, this interpretation raises broader questions. Here, we enumerate two 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 in markups accelerated in 1980 due to growing inequality. In Appendix Figure B13, we compile historical estimates from Barger (1955) and digitized copies of the 1969–1977 Census Annual Retail Trade Surveys. These estimates suggest markups were indeed lower in level and rising far before 1980 for several retail sectors. The historical time series should be interpreted with caution, since industry definitions, methods, and samples differ across these sources.

What about rising markups beyond retail? Our analysis focus on markups in retail, where we have rich data to quantify the forces relating markups to customer income. However, in some models of vertical supply chains, a decline in consumer price sensitivity can lead markups to rise along the entire producer chain (e.g., Tirole 1988 Ch. 3; Wu 2022). In Appendix F.1.1, we 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 customers.[29] These patterns suggest that the transmission of declining consumer price sensitivity to upstream firms merits further investigation.

9 Conclusion

Ideas relating the price elasticity facing firms to customer income trace back to Harrod's (1936) "Law of Diminishing Elasticity of Demand" and are supported by rich empirical

[29]Brand (2021) and Döpper et al. (2021) also find that markups set by upstream consumer good manufacturers depend on the income and price sensitivity of their ultimate consumers.

evidence in macroeconomics, trade, and industrial organization. In this paper, we show that income levels and income inequality are important determinants of markups in an economy. We provide new measures of how income affects aggregate markups and decompose this aggregate relationship into the effects of a household's own income and spillovers across households. In keeping with the empirical evidence, we develop a model in which the relationship between markups and aggregate income 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 due to how firms' prices respond to customer behavior.

The model suggests changes in the income distribution can play an important role in shaping markups across space and over time. Specifically, the model predicts that markups across cities rise with income levels and inequality, and explains empirical patterns of markups across cities with more success than standard supply-side models that predict markups based on firm market shares and concentration. The model can also account for a substantial share of the rise in retail markups over time and for reallocations of sales to high-markup firms. Altogether, these findings propose a demand-based explanation for rising markups and caution policy interpretations that attribute increasing markups solely to supply-side forces.

References

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

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

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

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

Albrecht, J., G. Menzio, and S. Vroman (2023). Vertical differentiation in frictional product markets. *Journal of Political Economy Macroeconomics 1*(3), 586–632. National Bureau of Economic Research, Working Paper No. 29618.

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. National Bureau of Economic Research, Working Paper No. 24434.

Anderson, E. T. and E. J. Fox (2019). How price promotions work: A review of practice and theory. *Handbook of the Economics of Marketing 1*, 497–552.

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.

Baker, S. R., B. Baugh, and M. C. Sammon (2023). Measuring customer churn and interconnectedness. *Journal of Political Economy Macroeconomics 1*(3), 447–505. National Bureau of Economic Research, Working Paper No. 27707.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

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.

Barsky, R. B., M. Bergen, S. Dutta, and D. Levy (2003). *Scanner data and price indexes*, Chapter What can the price gap between branded and private-label products tell us about markups?, pp. 165–228. University of Chicago Press.

Benkard, C. L., A. Yurukoglu, and A. L. Zhang (2021). Concentration in product markets. National Bureau of Economic Research, Working Paper No. 28745.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Bhatia, N., S. Moshary, and A. Tuchman (2021). Investigating the pink tax: Evidence against a systematic price premium for women in CPG. Technical report, Working paper.

Boar, C. and E. Giannone (2023). Consumption segregation. National Bureau of Economic Research, Working Paper No. 31133.

Bornstein, G. (2021). Entry and profits in an aging economy: The role of consumer inertia. 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.

Broda, C. and J. A. Parker (2014). The economic stimulus payments of 2008 and the aggregate demand for consumption. *Journal of Monetary Economics 68*, S20–S36.

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

Butters, R. A., D. W. Sacks, and B. Seo (2023). Racial difference in retail prices paid. Technical report, Working paper.

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.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *The Journal of Finance 63*(4), 1977–2011.

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

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

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

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

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

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

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

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

Grieco, P. L., C. Murry, and A. Yurukoglu (2024). The evolution of market power in the us auto industry. *The Quarterly Journal of Economics 139*(2), 1201–1253.

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. National Bureau of Economic Research, Working Paper No. 28136.

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

Jaimovich, N., S. Rebelo, and A. Wong (2019). Trading down and the business cycle. *Journal of Monetary Economics 102*, 96–121.

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.

Kesavan, S., B. R. Staats, and W. Gilland (2014). Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science 60*(8), 1884–1906.

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

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies 60*(3), 531–542.

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

Menzio, G. (2023). Optimal product design: Implications for competition and growth under declining search frictions. *Econometrica 91*(2), 605–639. National Bureau of Economic Research, Working Paper No. 28638.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory 18*(2), 301–317.

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.

Peltzman, S. (2000). Prices rise faster than they fall. *Journal of Political Economy 108*(3), 466–502.

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.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2021). Diverging trends in national and local concentration. *NBER Macroeconomics Annual 35*(1), 115–150.

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.

Smith, D. A. and S. Ocampo (2023). The evolution of U.S. retail concentration. Working paper.

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

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

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. (2023). The missing intercept: A demand equivalence approach. *American Economic Review 113*(8), 2232–2269. National Bureau of Economic Research, Working Paper No. 29558.

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

Online Appendix to *Markups Across the Income Distribution: Measurement and Implications*

Kunal Sangani

(Not for publication)

Appendix A Data Cleaning and Construction

A.1 NielsenIQ Homescan

Continuous measure of household income. NielsenIQ reports income in discrete categories. For analyses that use household income as a continuous variable, we follow Broda et al. (2009) and recode each household's income as the midpoint of the income bracket. For households with over $200,000 in annual income, we assign an income of $225,000. For years outside 2006–2009, incomes are topcoded at $100,000, and so for all analyses that span years beyond 2006–2009, we assign an income of $150,000 to all households with incomes over $100,000.

Treatment of retailer IDs. NielsenIQ provides a retailer ID for each transaction, though retailers' identities are anonymized. Large retail chains, even those that are not part of NielsenIQ Retail Scanner program, are given unique retailer IDs. Some retailer IDs capture multiple small retail chains. For analyses where we use retailer size as a control, we use the sales of the smallest uniquely-identified retailer as a proxy for the sales of small retail chains without unique identifiers. For the analysis removing the largest retailers from the sample (Figure B1), we rank retail chains by total sales in the Homescan data, assuming all retailers without unique identifiers are smaller than those identified.

A.2 PromoData Price-Trak

The PromoData include a list of active categories and inactive product categories. We use both the active and inactive databases and drop duplicated observations in the inactive database. Since UPCs may be available in multiple pack sizes, we call each unique UPC-pack size available to a retailer an "item" in the following description.

Data construction. We calculate monthly wholesale prices as the minimum reported base and deal price for each item in each market in each month. Consistent with Stroebel and Vavra (2019), we find that wholesale prices are surprisingly uniform across markets: Table A1 shows that over 80 percent of items in a given month have a wholesale cost exactly equal to the modal price across markets.[30] Aggregating across different items, we assume that retailers purchase UPCs at the minimum price available to them, and so we calculate wholesale base and deal prices for each UPC in each month by taking the

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

minimum price at which the UPC is offered across items (pack sizes) in that month. Since the PromoData lack information on the quantities sold, this is a more principled approach than taking an unweighted average across items.

Merge with NielsenIQ Homescan. We merge monthly wholesale costs into the Homescan data using the date of the shopping trip and the product UPC. Table A2 shows summary statistics on the distribution of sales-weighted average markups across UPCs.

Table A3 reports the share of purchases matched to wholesale costs by income group. The share of matched transactions (expenditures) increases (decreases) slightly with income. We compare relative prices for matched and unmatched transactions by income group. Relative prices are calculated as the percent difference in the unit price in a transaction relative to the average unit price for all transactions in the product module. The final two columns of Table A3 show that relative prices of unmatched products grow with income faster than relative prices of matched products, suggesting that markup differences in our merged sample are conservative. We use relative unit prices paid across income groups in Appendix G.2 to estimate the bias in the markup gap due to selection.

To check whether assuming uniformity of wholesale prices across markets materially affects the results, we replicate analyses by matching wholesale costs only to transactions made by households living in the same market as the wholesaler. We use a hand-constructed crosswalk from Scantrack Market IDs in the NielsenIQ Homescan panel to PromoData Price-Trak market areas. The sample matched at the wholesaler market level includes 3.0 million transactions (1.8 million at stores with unique NielsenIQ IDs). The analyses replicated using this subset of the data are reported in Table 1.

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 in bulletin 20-01, which we use to map county FIPS codes to CBSAs throughout.
- **CBSA income distribution and inequality measures.** We use the number of households by income bucket (variable BE19001) and Gini indices (B19083) by CBSA from the 2009–2013 American Community Survey 5-year estimates.
- **Retail grocery establishments.** We use annual data on the number of NAICS 445 establishments (includes grocery stores, supermarkets, liquor stores, and specialty food stores) by county from the Census County Business Patterns.

Table A1: Uniformity of wholesale prices across markets.

	Measure of wholesale cost	
	Deal Price	Base Price
Percent of items listed:		
At modal price	78.5	80.3
Within 5% of modal price	86.4	90.7
Within 10% of modal price	90.9	95.1

Table A2: Summary statistics for markup distribution.

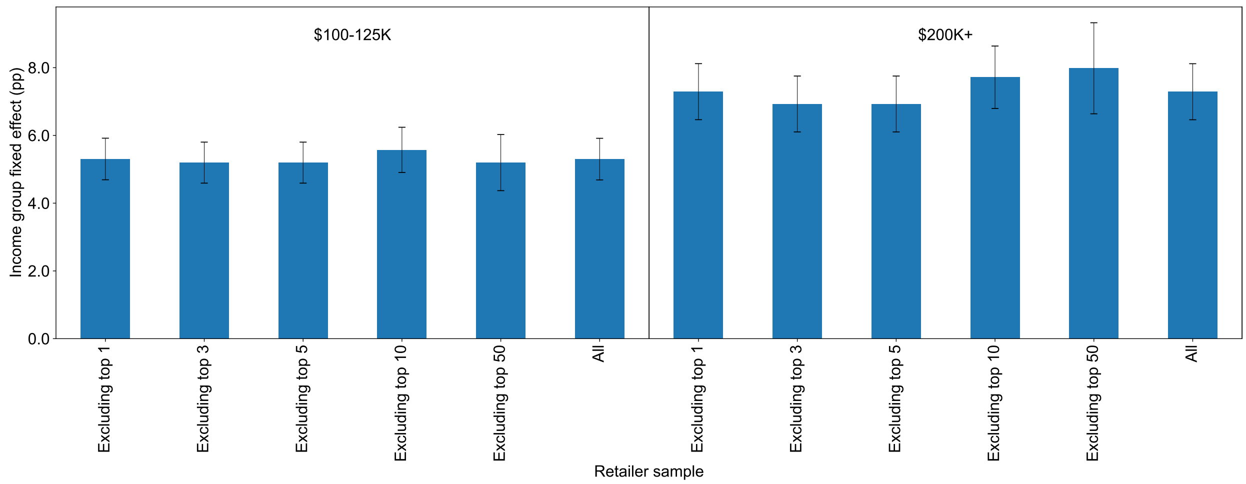
	Measure of wholesale cost	
	Deal Price	Base Price
Percentiles of distribution:		
10	1.119	1.053
25	1.288	1.204
50	1.470	1.382
75	1.694	1.600
90	2.002	1.911
Percent below $\mu = 1$:		
By count	4.72	6.96
By sales	6.35	12.63

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

	Percent matched to wholesale cost data		Log unit price, relative to product module average	
Income group	Transactions	Expenditures	Matched	Unmatched
\$10–25K	41	38	-0.02	-0.05
\$25–40K	42	38	0.00	-0.02
\$40–60K	43	38	0.04	0.02
\$60–100K	44	37	0.09	0.09
Over \$100K	44	35	0.17	0.17
All	43	37	0.06	0.05

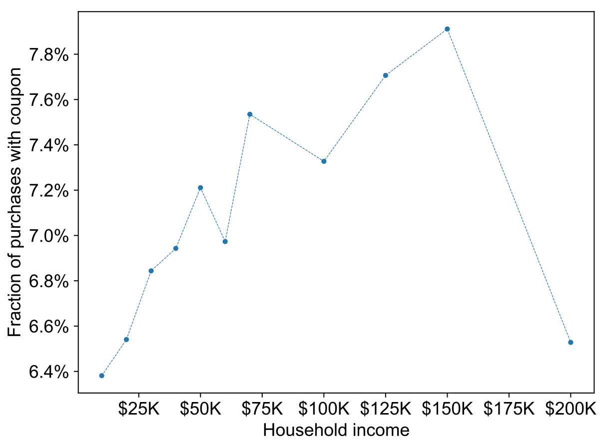
Appendix B Additional Tables and Figures

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

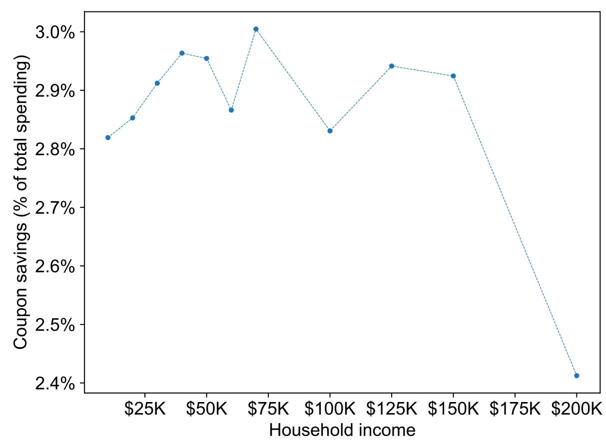


Note: The figure shows the fixed effect estimated in (1) for two high-income groups as large retailers (ranked by total sales in the Homescan data) are removed from the sample.

Figure B2: Coupon usage and savings by income.

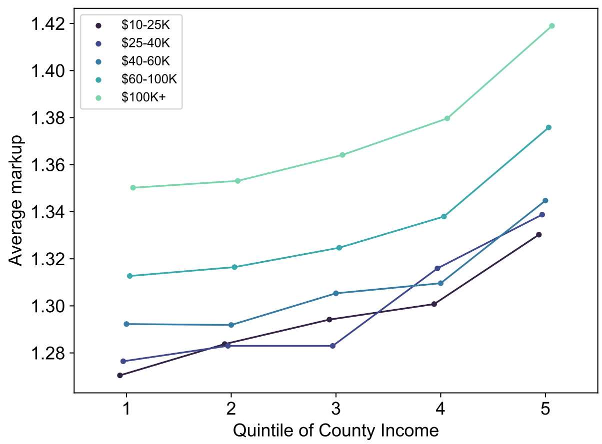


(a) Fraction of purchases made with a coupon.

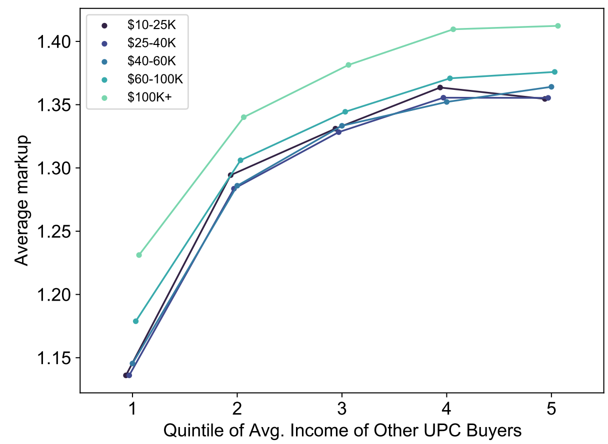


(b) Savings due to coupon use.

Figure B3: Average retail markup by income group, split by other buyers' incomes.

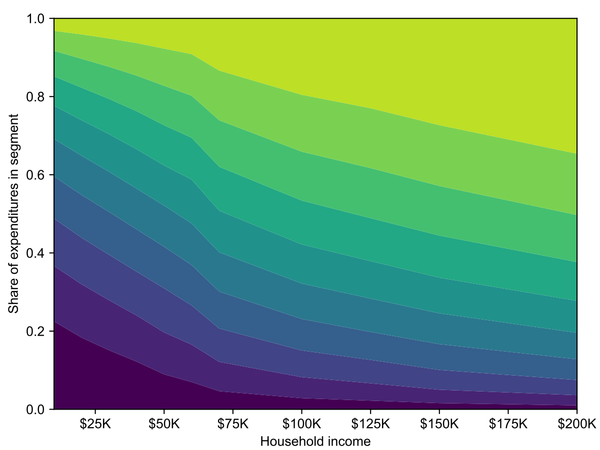


(a) County income

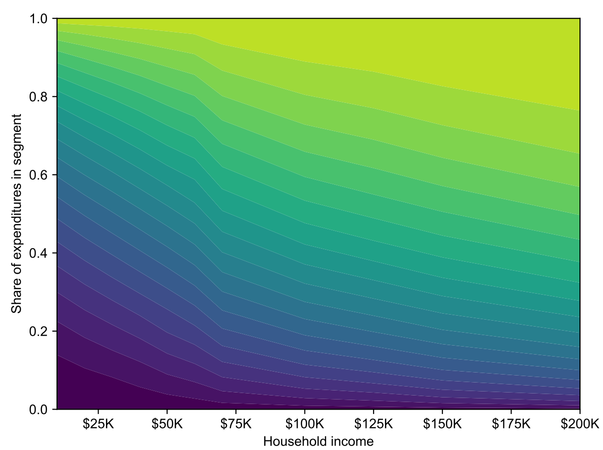


(b) Income of UPC's other buyers

Figure B4: Expenditure shares across $K = 10$ and $K = 20$ groups of UPCs sorted by average buyer income.

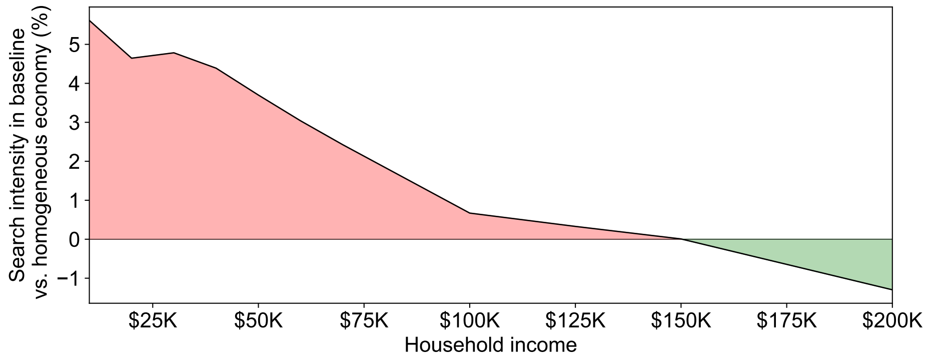


(a) $K = 10$ groups.

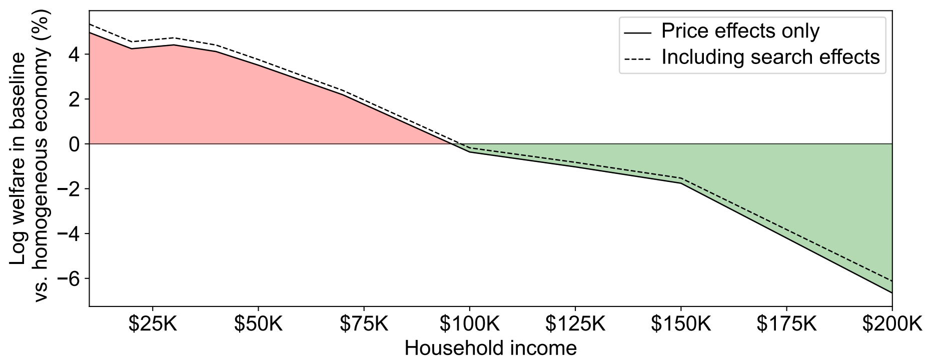


(b) $K = 20$ groups.

Figure B5: Spillovers of income level on search intensity and welfare.



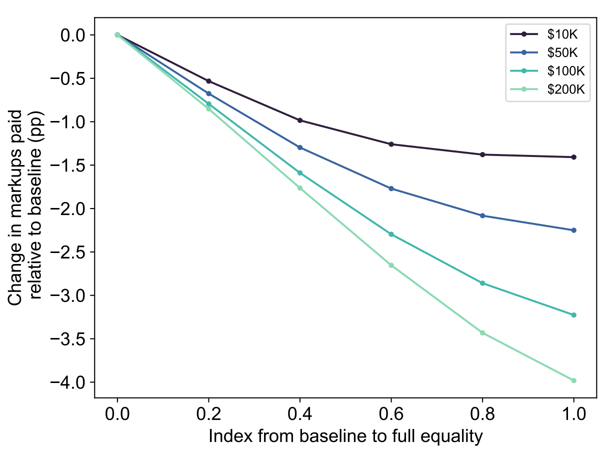
(a) Price quotes per purchase relative to homogeneous economies.



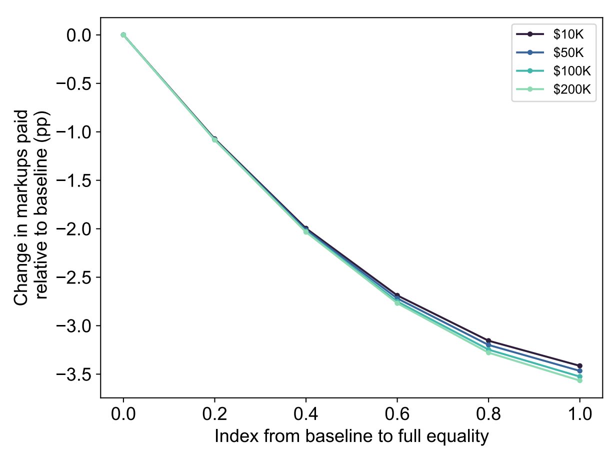
(b) Welfare relative to homogeneous economies.

Note: Welfare effects include effects from price changes (price effects only) and from changes in income due to changes in search time. For top income groups, search effects increase welfare since higher prices reduce consumption and thus overall search time, even though search per unit of consumption increases.

Figure B6: Effects of income inequality on markups in baseline and single-good models.

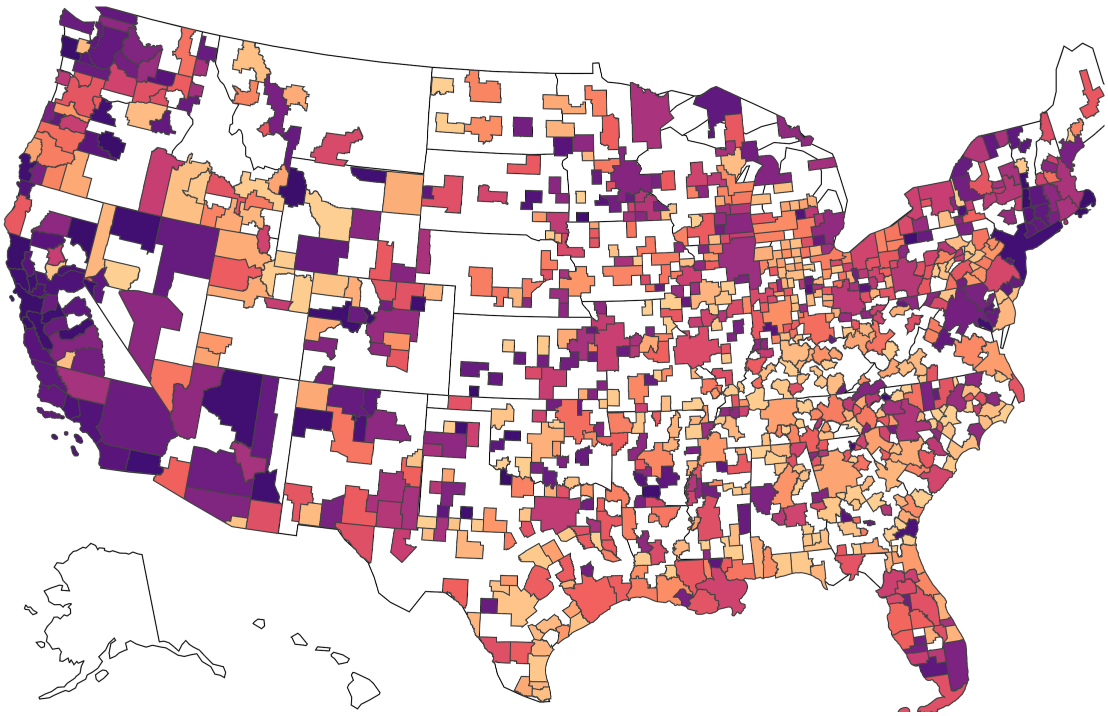


(a) Baseline model ($K = 10$).



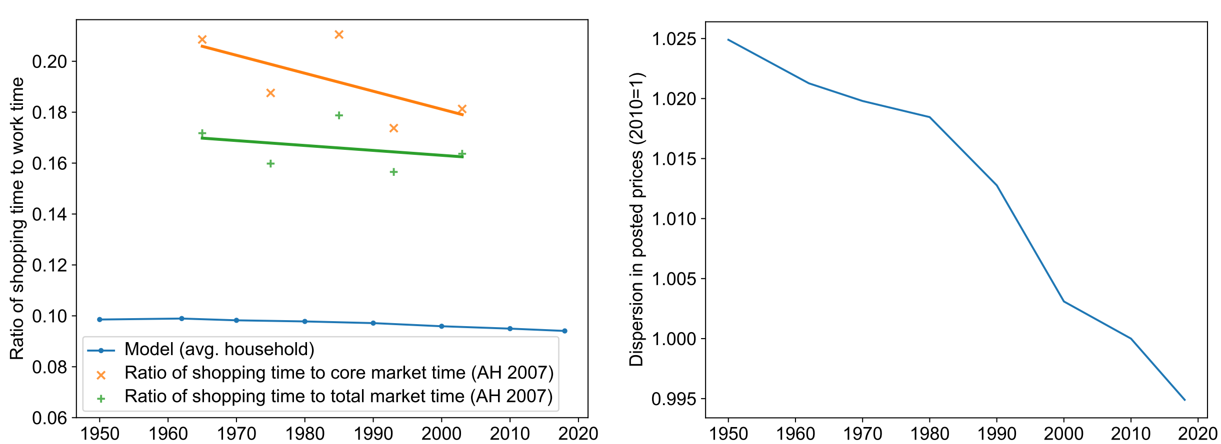
(b) Single-good model ($K = 1$).

Figure B7: Markups across CBSAs in the data.



Note: CBSAs are colored by rank of aggregate markup 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).

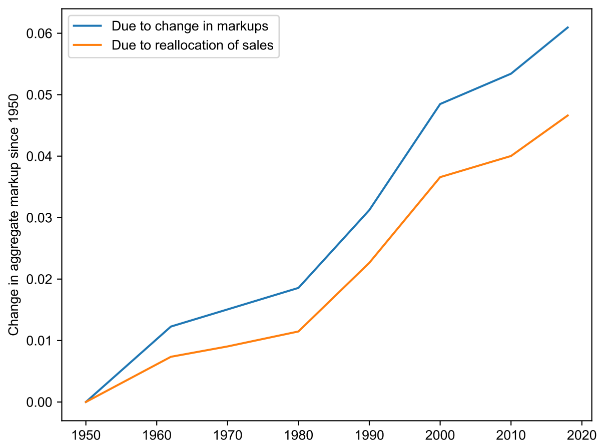
Figure B8: Shopping time and price dispersion in the model.



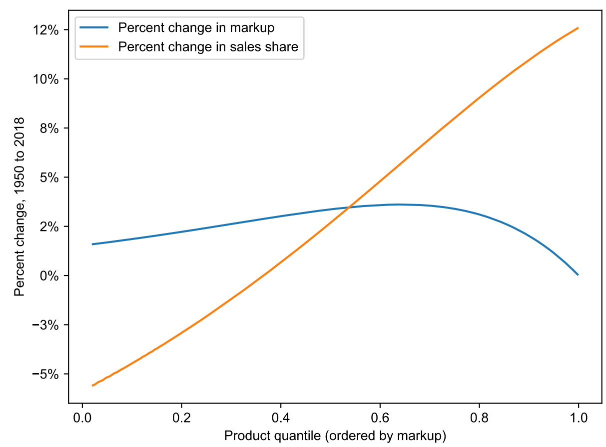
(a) Ratio of total shopping time to work time.

(b) Dispersion in posted prices.

Figure B9: Role of within-firm markup changes and reallocations across firms.

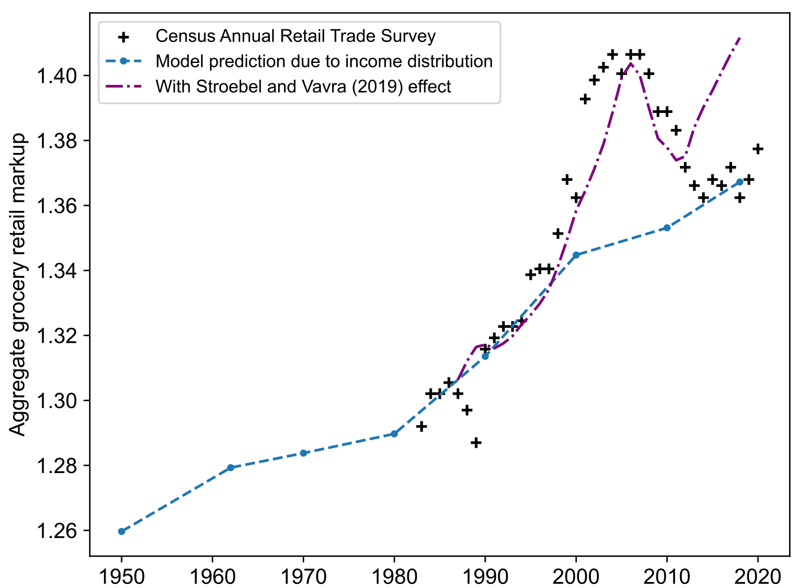


(a) Decomposition of change in agg. markup.



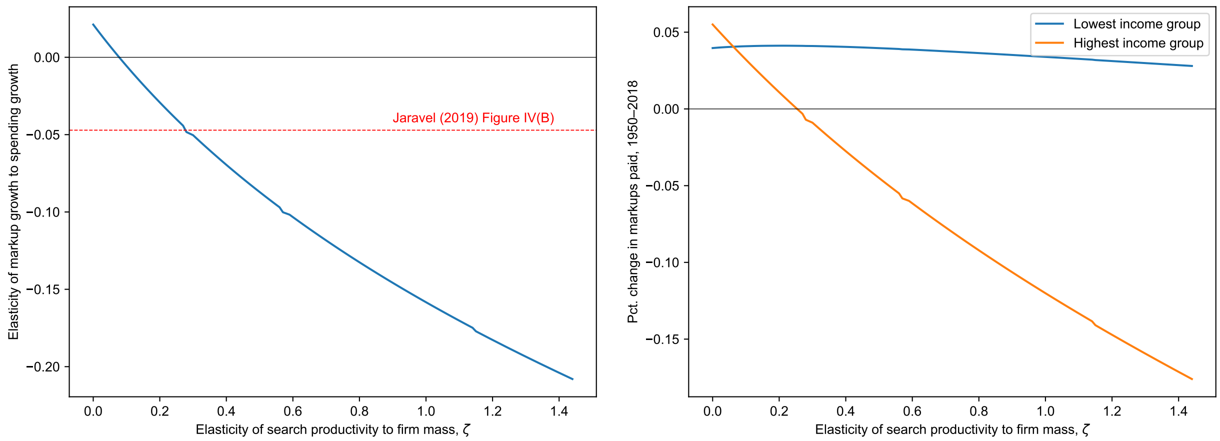
(b) Percent change in markups and sales shares by quantile for $k = 1$, 1950 to 2018.

Figure B10: 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. We apply an elasticity of markups to house prices of 7 percent (midpoint of the range of OLS estimates from Stroebel and Vavra 2019) to changes in the S&P / Case-Shiller national home price index since 1987.

Figure B11: Calibrating the pro-competitive effect parameter, ζ.

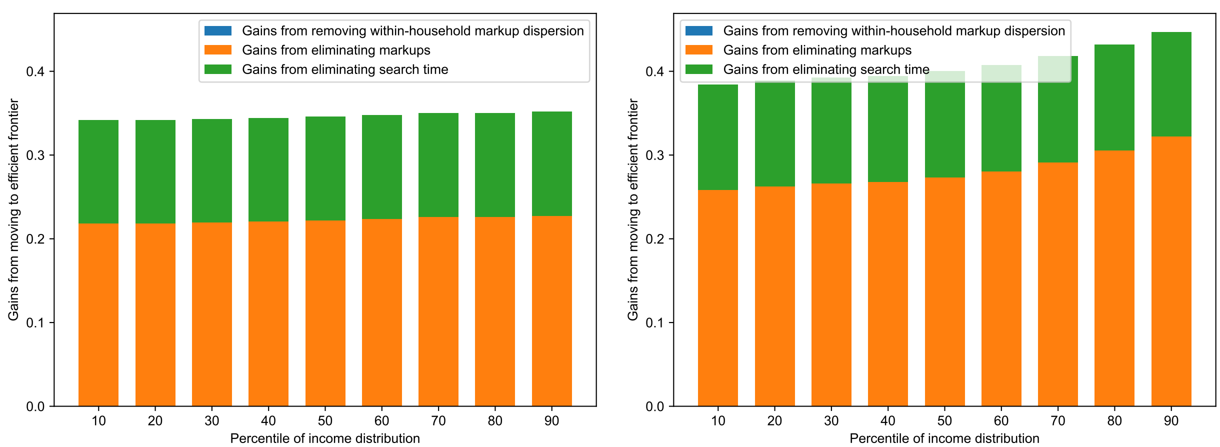


(a) Elasticity of markup growth to spending growth across $K = 10$ goods, 1950 to 2018.

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

Note: The left panel plots the coefficient estimated in 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 B12: Gains from moving to the efficient frontier across income percentiles.

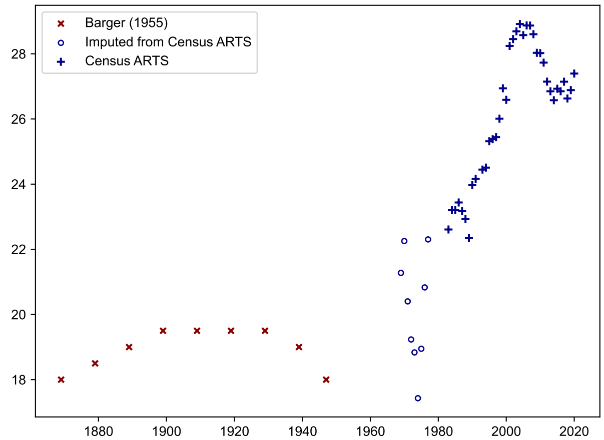


(a) 1950.

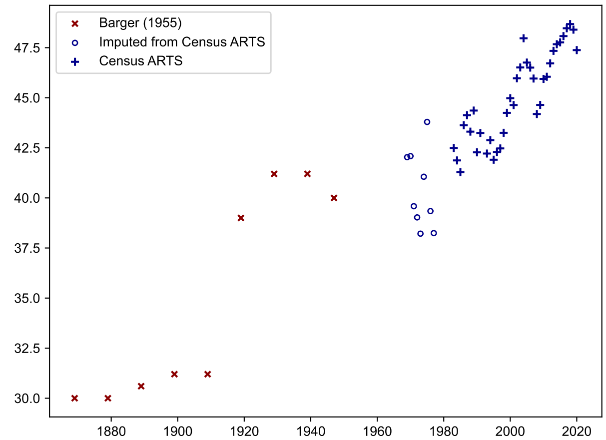
(b) 2018.

Note: The gains from eliminating within-household markup dispersion (blue) are welfare gains from setting households' markups across all goods equal to the markup they pay in aggregate. The gains from eliminating markups (orange) are welfare gains change from setting markups equal to one, holding search time constant. The gains from eliminating search time (green) are welfare gains from reallocating all search time to labor market time. Note that the gains from eliminating within-household markup dispersion (blue) are very small compared to the other two components.

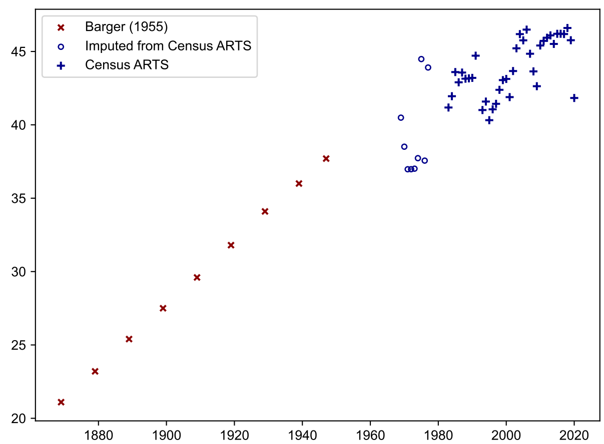
Figure B13: Data on retail gross margins over time by subsector.



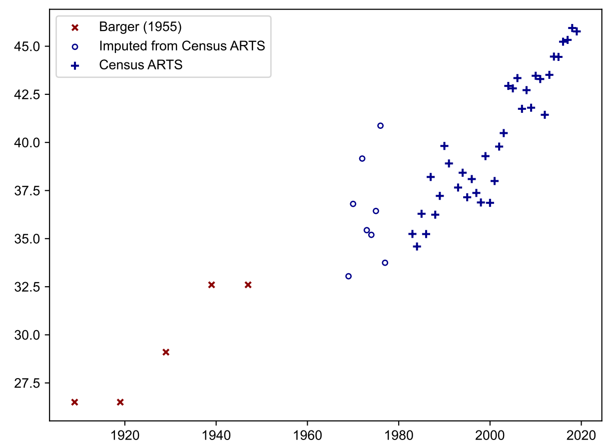
(a) Grocery stores



(b) Furniture stores



(c) Apparel stores



(d) Automobile accessory stores

Note: Gross margin estimates are available for selected years from 1869 to 1947 from Barger (1955), and annually from the Census Annual Retail Trade Survey from 1983 to 2020. Additionally, gross margins can be imputed from annual data on sales, purchases, and changes in inventories from the Census Annual Retail Trade Survey from 1969 to 1977. Gross margins are reported as total sales less total costs of goods sold as a percent of sales. For ARTS estimates, grocery stores include SIC 541 until 1992 and NAICS 4451 after 1993. Furniture stores include SIC 571 until 1992 and NAICS 442 after 1993. Apparel stores include SIC 56 until 1992 and NAICS 448 after 1993. Automobile parts and accessory stores include SIC 553 until 1992 and NAICS 4413 after 1993.

Table B1: Elasticity of markups paid for identical products to income, compared to overall elasticity of markups to income.

Log Retail Markup	Elasticity Within Identical Products Only (1)	(2)	Elasticity Including Across Products (3)	(4)
Log Household Income	0.009**	0.014**	0.022**	0.034**
	(0.001)	(0.001)	(0.002)	(0.004)
Demographic Controls	Yes	Yes	Yes	Yes
Store-UPC FEs	Yes			
County-UPC FEs		Yes		
Store FEs			Yes	
County FEs				Yes
Retailer–Product Controls		Yes		Yes
N (millions)	14.0	25.8	14.0	25.8
R^2	0.91	0.79	0.11	0.06

Note: Columns 1 and 2 report results from the specifications,

$$\log \text{Markup}_{igk} = \beta \log \text{Income}_i + \gamma' X_i + \alpha_{s(k),g} + \epsilon_{igk},$$
$$\log \text{Markup}_{igk} = \beta \log \text{Income}_i + \gamma' X_i + \delta' W_k + \phi_{c(k),g} + \epsilon_{igk},$$

where Markup_{igk} is the markup paid by household i on UPC g in transaction k, $\alpha_{s(k),g}$ are store-UPC fixed effects, and $\phi_{c(k),g}$ are county-UPC fixed effects, and the remaining variables are as defined in equations (1a) and (2a). Columns 3 and 4 report results from specifications (1a) and (2a). Regressions weighted by costs, and standard errors two-way clustered by product brand and household county.

Table B2: Impact of own and other buyers' incomes on markups paid in the cross section.

Log Retail Markup	(1)	(2)	(3)
Log Household Income	0.038**	0.026**	0.016**
	(0.005)	(0.004)	(0.001)
Log Avg. CBSA Income	0.083**		
	(0.011)		
Log Income at Retailer Locations		0.166**	
		(0.038)	
Log UPC Avg Income			0.211**
			(0.042)
Demographic Controls	Yes	Yes	Yes
Retailer–Product Controls	Yes	Yes	Yes
Store County FEs		Yes	
Store FEs			Yes
N (millions)	23.8	9.0	14.0
R^2	0.04	0.04	0.12

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

Table B3: Cross section vs. time series elasticities of markups to aggregate income.

Outcome	*Log Markup*		*Log Avg. Unit Price*			
Sample	2006–2012		2004–2019		2004 and 2019	
	(1)	(2)	(3)	(4)	(5)	(6)
Log CBSA Income	0.104**	0.099**	0.127**	0.128**	0.097**	0.096*
	(0.017)	(0.026)	(0.026)	(0.030)	(0.045)	(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
N (millions)	133	133	18.3	18.3	2.2	2.2
R^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%.

Table B4: Search time increases with basket size: Cross-sectional evidence.

	Log Shopping Trips		*Log Unique Stores*		*Log Unique Retailers*	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Log Expenditures	0.588**	0.617**	0.176**	0.044**	0.432**	0.159**
	(0.004)	(0.013)	(0.004)	(0.012)	(0.004)	(0.013)
Log Markup Paid	−0.688**	−0.694**	−0.172**	−0.144**	−0.344**	−0.285**
	(0.025)	(0.025)	(0.028)	(0.028)	(0.025)	(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 Markup Paid is the aggregate markup paid by the household over all observed purchases. Demographic controls include race, ethnicity, and presence and age 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 B5: Search time increases with basket size: Time series evidence.

	Log Shopping Trips		*Log Unique Stores*		*Log Unique Retailers*	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Log Expenditures	0.745**	0.143**	0.527**	0.203**	0.193**	0.115**
	(0.010)	(0.029)	(0.006)	(0.032)	(0.002)	(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 B6: Elasticity of markups paid to CBSA income is larger for high-income households, both in cross-section and time series.

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

Note: Demographic controls include race, ethnicity, and presence and age group of female head of household. Mid-Income Group is an indicator equal to one for households with incomes between $50K and $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 B7: Robustness: Counterfactual on rise in markups, 1950–2018.

Alternate parameter values:	Predicted Δ in markup	Due to Δ Income level	Due to Δ Income dispersion	Due to Change in markups	Due to Reallocations of sales
Baseline	10.8	8.3	2.5	6.1	4.7
Number of products:					
$K = 1$	14.3	11.3	3.0	10.1	4.2
$K = 3$	11.3	8.7	2.6	6.7	4.6
$K = 5$	11.0	8.4	2.5	6.5	4.5
$K = 20$	10.7	8.2	2.4	6.0	4.6
$K = 50$	10.6	8.2	2.4	5.9	4.7
$K = 100$	10.6	8.2	2.4	5.9	4.7
Reservation price:					
$R = 2.8$	10.8	8.3	2.5	6.2	4.6
$R = 3.3$	10.5	8.1	2.4	5.8	4.6
$R = 3.5$	10.3	8.0	2.3	5.8	4.5
Elasticity of substitution:					
$\sigma = 0$	10.8	8.3	2.5	6.1	4.7
$\sigma = 10$	10.8	8.3	2.5	6.4	4.4
Pro-competitive effects:					
$K = 1, \zeta = 0.3$	11.5	9.2	2.3	8.6	2.9
$K = 10, \zeta = 0.3$	9.5	7.6	1.9	5.5	4.0
Alternative expenditure shares:					
NielsenIQ-tracked categories	10.2	8.1	2.1	5.7	4.5
Food-at-home, housekeeping supplies, and personal care	10.1	8.1	2.0	5.7	4.4
Food-at-home	9.9	8.0	1.9	5.6	4.3
All food	10.3	8.2	2.2	5.9	4.5

Note: Alternative expenditure shares use expenditure shares by income from the 2007 Consumer Expenditure Survey (Table 55). NielsenIQ-tracked categories include food-at-home, alcoholic beverages, housekeeping supplies, small appliances and miscellaneous housewares, and personal care.

Table B8: Evolution of inequality in costs of goods purchased in the model over time.

Gini indices	Baseline		1950	2018	Change	
Post-tax income	46.6	–	34.0	48.7	+14.7	–
Costs of goods purchased (baseline, $\zeta = 0$)	45.7	*-2.0%*	33.6	47.8	+14.2	*-3.6%*
Incl. pro-competitive effects ($\zeta = 0.3$)	45.7	*-2.0%*	33.6	47.8	+14.1	*-4.1%*

Appendix C Proofs

C.1 Households

Let $p(s, F)$ denote the average price paid for a good with price distribution F given search intensity s. Note that $p(s, F)$ can be written as

$$p(s, F) = \sum_{n=1}^{\infty} q_n(s)\mathbb{E}[p|n], \tag{12}$$

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

$$\mathbb{E}[p|n] = \int_{\underline{p}}^{R} pd\left[1 - (1 - F(p))^n\right] = \int_{\underline{p}}^{R} pn\left(1 - F(p)\right)^{n-1} dF(p) = \underline{p} + \int_{\underline{p}}^{R} (1 - F(p))^n dp. \tag{13}$$

Lemma 3 (Expected price with n quotes). *For any nondegenerate 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 first part of the lemma is immediate from derivatives of (13) with respect to n. 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 $\{Q_n\}_{n=0}^{\infty}$ to rewrite (12) 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 take the first and second derivatives of the price function with respect to search effort s, which we denote p_s and p_{ss}:

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

Lemma 4 (Properties of the price function). *Under Assumption 1, for any nondegenerate price distribution F, $p_s < 0$ and $p_{ss} > 0$.*

Proof. From Lemma 3, $\mathbb{E}[p|n] - \mathbb{E}[p|n+1] > 0$, 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 $p_{ss} > 0$. ■

Recall the households' maximization problem:

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

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, which we can write more simply (and without subscripts) as $\phi = -p_s(s,F)$. By Lemma 4, $-p_s$ is strictly decreasing in s, so we can invert this equation to write $s = s(\phi, F)$. It will be helpful to note:

$$\frac{\partial s}{\partial \phi} = \frac{1}{-p_{ss}} < 0, \qquad \text{and} \qquad \frac{\partial^2 s}{\partial \phi^2} = \frac{p_{sss}}{-(p_{ss})^3}. \tag{14}$$

Proof of Lemma 1. Using (14), 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 3 that $p_s < 0$. ■

C.2 Firms

For convenience, we 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, and so on. Hence, the demand curve facing a firm charging price $p \leq R$ is

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

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

$$\pi(p) = \frac{C}{M}(p-1)\sum_{n=1}^{\infty} n\bar{q}_n(1-F(p))^{n-1}. \tag{15}$$

Our equilibrium condition for $F(p)$ 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

$$\pi = \frac{C}{M}(R - 1)\,\bar{q}_1. \tag{16}$$

Setting profits in (15) equal to (16) and solving yields the expressions for F and the minimum price $\underline{p}$ in (10).

Proof of Lemma 2. Total sales over total variable costs for good k, $\bar{\mu}_k$, is:

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

The aggregate markup is the cost-weighted average across the K products. ∎

C.3 Equilibrium stability and strategic interactions

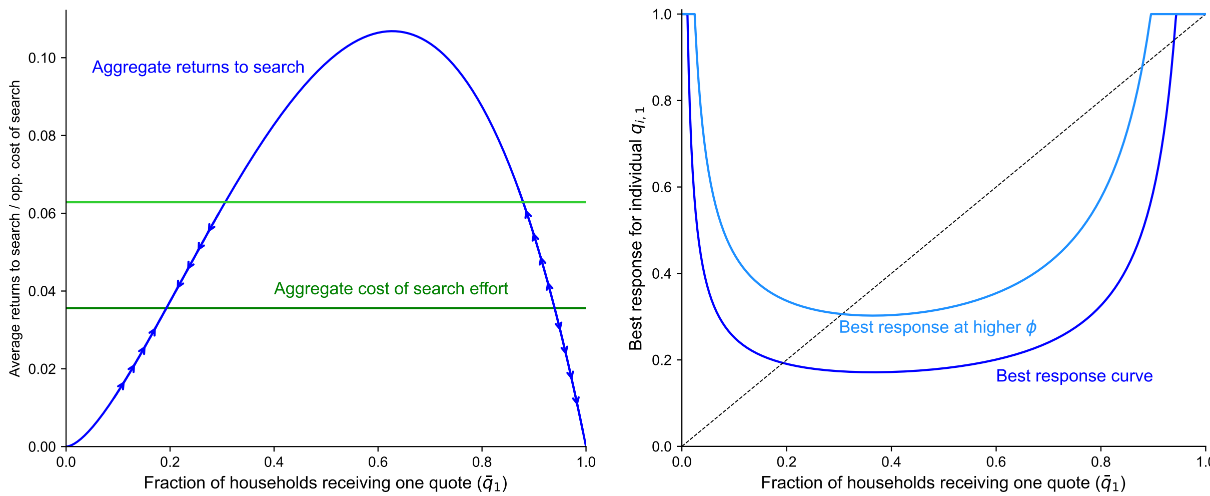
Aggregating households' first order conditions in (7) yields the equilibrium condition,

$$\underbrace{\int_0^\infty -p_s(s(z), F)d\Lambda(z)}_{\text{Aggregate returns to search}} = \underbrace{\int_0^\infty \phi(z)d\Lambda(z).}_{\text{Aggregate cost of search effort}} \tag{17}$$

The left panel of Figure C1 illustrates this equilibrium intersection and how aggregate returns to search vary with the share of households receiving only one quote, $\bar{q}_1$. Notice that there are generally two dispersed-price equilibria. As indicated by the arrows along the curve for returns to search, the left-most equilibrium is stable, while the right equilibrium is unstable, since changes in any one household's search intensity at this equilibria kick off changes across all households that lead away from the equilibrium point.

Lemma 5 shows that, when search costs are sufficiently low, household search decisions are strategic substitutes in equilibrium.

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



(a) Stable dispersed-price equilibrium. **(b)** Best response of individual household.

Lemma 5 (Strategic substitutes in search)**.** *There exists some ϕ^{cutoff} such that, if $\phi(z) < \phi^{cutoff}$ for all z, then (1) a 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).*

Proof of Lemma 5. Let us rewrite the price paid for by household i directly in terms of the search intensities chosen by i and all other households: $\hat{p}\left(s_i, \{s_k\}\right) \equiv p(s_i, F)$. The household's first order condition is thus $-\hat{p}_s\left(s_i, \{s_k\}\right) = \phi_i$. Taking the implicit derivative relative to a change in some other household j's search behavior yields

$$\frac{\partial s_i}{\partial s_j} = \frac{-\hat{p}_{s,s_j}}{\hat{p}_{ss}}.$$

From Assumption 1, $\hat{p}_{ss} > 0$ (Lemma 4). The numerator $-\hat{p}_{s,s_j}$ has a natural interpretation as how returns to search change when aggregate search intensity increases due to an increase in search intensity by one household.

At the limit where $\phi_i \to 0$ for all i, $\bar{q}_1 \to 0$, $F(p)$ approaches a degenerate distribution with point mass at $p = 1$, and since F is degenerate, returns to search $-p_s = 0$. Now suppose $\phi_j = \varepsilon > 0$ for some j. Since this is an increase in the cost of search for j, this perturbation results in a decrease in s_j, 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 $-p_{s,s_j} < 0$. Hence, we can conclude that there exists some ϕ^{cutoff} such that if $\phi_i < \phi^{\text{cutoff}}$ for all i, then $\partial s_i/\partial s_j < 0$.

It remains to show that a dispersed-price equilibrium exists. This follows from the fact that we can choose ϕ^{cutoff} low enough to guarantee that the aggregate cost of search effort in (17) is low enough and the equilibrium condition (17) holds.[31] ∎

The right panel of Figure C1 illustrates a case where costs of search are low enough that search decisions are strategic substitutes. This can be seen by noting that the best response curve for an individual household is downward-sloping when it intersects the 45-degree line. Hence, a marginal increase in the fraction of households receiving one quote $\bar{q}_1$ leads to a decline in the household's best response for $q_{i,1}$.

C.4 Example with identical households

Proof of Proposition 1. Suppose households are initially identical. If households are identical, markets for each good are also identical. Thus, prices paid by any household i can be summarized by

$$p_i = p(s_i, F(s_{-i})) \equiv \hat{p}(s_i, s_{-i}),$$

where we define the function $\hat{p}$ that takes household i's search intensity s_i and the aggregate search intensity s_{-i} directly as arguments. Taking the derivative with respect to an infinitesimal change in household i's labor productivity, dz_i, and change in the labor productivity of other households, dz_{-i}, yields

$$dp_i = \hat{p}_s \left[\frac{\partial s_i}{\partial z_i} dz_i + \frac{\partial s_i}{\partial s_{-i}} \frac{ds_{-i}}{dz_{-i}} dz_{-i} \right] + \hat{p}_{s_{-i}} \frac{ds_{-i}}{dz_{-i}} dz_{-i},$$

where $\hat{p}_s$ and $\hat{p}_{s_{-i}}$ denote the partial derivatives of $\hat{p}$ with respect to s_i and s_{-i}.

Recall that each household's first order condition is $-p_s = z_i/a_i$. Taking the total derivative yields

$$-\hat{p}_{ss} ds_i - \hat{p}_{ss_{-i}} ds_{-i} = \frac{1}{a_i} \left(1 - \frac{z_i}{a_i} \frac{da_i}{dz_i} \right) dz_i,$$

which means

$$\frac{\partial s_i}{\partial z_i} = -\frac{1}{a_i} \frac{1}{\hat{p}_{ss}} \left(1 - \frac{d \log a_i}{d \log z_i} \right), \qquad \text{and} \qquad \frac{\partial s_{-i}}{\partial z_{-i}} = -\frac{1}{a_i} \frac{1}{\hat{p}_{ss} + \hat{p}_{ss_{-i}}} \left(1 - \frac{d \log a_i}{d \log z_i} \right).$$

Plugging these expressions in, and using the fact that the marginal cost is equal to the

[31] When search costs are too large, households choose not to search, and the dispersed-price equilibrium ceases to exist. There are intermediate values of ϕ for which the dispersed-price equilibrium exists and household search decisions are strategic complements. However, the strategic substitutes behavior in Lemma 5 is the relevant region both in the calibrated model and in data on search behavior.

numeraire, we find

$$d\mu_i = \left(1 - \frac{d\log a}{d\log z}\right)\left[\kappa_1 dz_i + \kappa_2 \frac{\partial s_i}{\partial s_{-i}} dz_{-i} + \kappa_3 dz_{-i}\right],$$

where

$$\kappa_1 = -\frac{1}{a}\frac{\hat{p}_s}{\hat{p}_{ss}}, \qquad \kappa_2 = -\frac{1}{a}\frac{\hat{p}_s}{\hat{p}_{ss} + \hat{p}_{ss_{-i}}}, \qquad \text{and} \qquad \kappa_3 = -\frac{1}{a}\frac{\hat{p}_{s_{-i}}}{\hat{p}_{ss} + \hat{p}_{ss_{-i}}}.$$

Lemma 4 establishes that, under Assumption 1, $\hat{p}_s < 0$ and $\hat{p}_{ss} > 0$. An increase in s_{-i} leads to a first-order stochastic shift in $\bar{q}$ and hence an decrease in $\mathbb{E}[p|n]$ for all n. Thus, by (12), $\hat{p}_{-s} < 0$. Finally, the stability of the equilibrium guarantees that $(\hat{p}_{ss} + \hat{p}_{ss_{-i}}) > 0$. To see this, we can write household adjustments in search behavior around the equilibrium in terms of the dynamic system,

$$\frac{ds_i}{dt} = \alpha\left(-\hat{p}_s\left(s_i, s_{-i}\right) - \phi_i\right),$$

where ds_i/dt describes how households adjust search behavior (households increase search effort when the returns to search are greater than the opportunity cost of search effort, $-\hat{p}_s > \phi_i$) and α describes the speed of convergence. Taking the fact that households are uniform (i.e., $s_i = s_{-i}$) and denoting $ds/dt = g(s)$, the system is stable if $g'(s) < 0$. Thus, stability implies:

$$-\alpha\left(\hat{p}_{ss} + \hat{p}_{s,s_{-i}}\right) < 0,$$

and $(\hat{p}_{ss} + \hat{p}_{ss_{-i}}) > 0$. Plugging these in, we verify that $\kappa_1, \kappa_2, \kappa_3 > 0$. ■

C.5 Comparative statics

Recall from Lemma 2 that the aggregate markup when $K = 1$ is

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

Thus, to characterize the response of $\bar{\mu}$ to a perturbation in $\Lambda(z)$, we need to characterize the response of $\bar{Q}_1$.

Proof of Proposition 2. 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 $\bar{\mu}$. ■

Proof of Proposition 3. 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

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

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

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

Rearranging yields,

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

which is exactly the condition guaranteed by Assumption 2. ■

These comparative statics are illustrated in the left panel of Figure C1. An increase in the aggregate opportunity cost of search effort, $\int \phi(z) d\Lambda(z)$, increases the fraction of households receiving one quote $\bar{q}_1$ in equilibrium, thus raising the aggregate markup.

C.6 Application to two-quote and Poisson cases

We show that Assumptions 1 and 2 both hold under two common parameterizations of the search mapping function $\mathcal{S}$: a two-quote case and the Poisson case.

Two-quote. 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 $q_{i,2} = 1 - \exp(-s_i)$.

Assumption 1 becomes $\exp(-s_i)\left[\mathbb{E}\left[p|1\right] - \mathbb{E}\left[p|2\right]\right] > 0$, which holds since $\exp(-s_i) > 0$ and $\mathbb{E}\left[p|n\right]$ is strictly decreasing in n. Assumption 2 becomes:

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

So, we verify that the two-quote mapping satisfies both Assumption 1 and Assumption 2.

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

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

Dropping the i subscripts for convenience, some algebra yields:

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

We can verify that both Assumption 1 and Assumption 2 hold:

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

Appendix D Bias from Unobserved Local Costs

If some labor and rent costs are in fact variable, these unobserved local costs can bias the elasticity of markups to aggregate income measured across space. In this appendix, we gauge the potential magnitude of this bias using data on retail operating expenses and data on retail wages and rents.

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. We 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.[32]

The aggregate markup constructed from cost of goods sold alone μ^{COGS} and the "true" aggregate markup μ^{true} are given by $\mu^{\text{COGS}} = pY/cY$ and $\mu^{\text{true}} = pY/VC(Y)$, where pY are total sales. Taking the elasticity with respect to aggregate income I yields:

$$\frac{d \log \mu^{\text{COGS}}}{d \log I} = \underbrace{\frac{d \log \mu^{\text{true}}}{d \log I}}_{\text{True elasticity}} + \underbrace{\frac{wL}{VC} \frac{d \log w}{d \log I} + \frac{rA}{VC} \frac{d \log r}{d \log I}}_{\text{Bias}}. \tag{18}$$

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 (18) disappears.

Table D1 lists costs of goods sold, labor expenses, and rent expenses for four retail sectors from the 2007 Census Annual Retail Trade Survey of Detailed Operating Expenses. Even with expansive definitions that include all payroll, fringe benefits, commissions, and contractor costs for labor, and lease and repair expenses for both stores and offices/other buildings for rents, these are small portions of plausibly variable costs (15 and 3 percent).

We estimate the elasticities of factor costs to local income using retail wages from the Bureau of Labor Statistics Occupational Employment and Wage Statistics (OEWS) and retail rents from Moody's REIS platform; these estimates are reported in Table D2.

[32]The use of labor in proportion to output likely falls with income, making the below estimates conservative. Using data from the BLS Current Employment Statistics, we find that per-capita employment in retail trade has an elasticity of 0.24 to CBSA per-capita income. The elasticity of per-capita retail receipts to CBSA per-capita income is 0.58 (Census Statistics of U.S. Businesses). Thus, even with an elasticity of markups to income ≈ 0.11, the quantity of retail labor used relative to retail output appears to fall with income.

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
Total operating expenses	873,400	736,072	128,985	116,175
Total labor expenses	468,089	387,610	74,663	68,375
Total rent expenses	90,391	79,666	11,647	9,806
Labor cost / (OPEX + COGS) (%)	12.4	13.5	14.4	14.7
Rent cost / (OPEX + COGS) (%)	2.4	2.8	2.3	2.1

Note: Sales and gross margins from the 2007 Census Annual Retail Trade Survey, and operating expenses from the 2007 Census Retail Trade Survey of Detailed Operating Expenses. Labor expenses include payroll, fringe benefits, contract labor costs, and commission expense. Rent expenses include lease/rental payments for stores/offices and repairs and maintenance for stores/offices.

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

	Log Retail Wages (OEWS)		*Log Retail Rents (REIS)*	
Variable:	Cashiers (1)	Retail Salespersons (2)	Asking Rent (3)	Effective Rent (4)
Log Avg. CBSA Income	0.285**	0.265**	1.226**	1.265**
	(0.040)	(0.027)	(0.203)	(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. All estimates use data from 2007. Robust standard errors in parentheses. ** indicates significance at 5%.

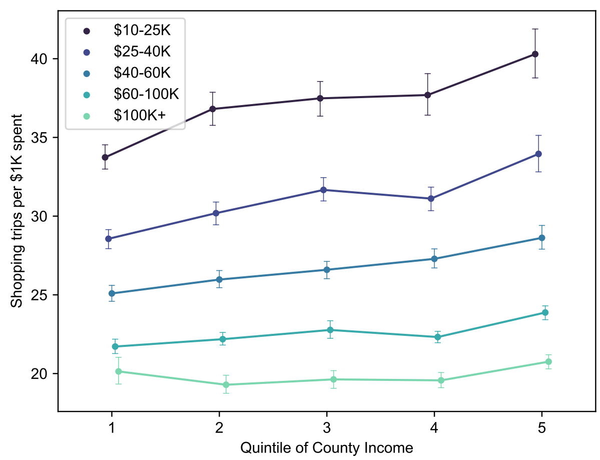
When all labor and rent costs are fixed, $d\log\mu^{\text{true}}/d\log I = d\log\mu^{\text{COGS}}/d\log I \approx 11.0\%$ (from Table 5). At the extreme where all labor and rent costs are variable, the elasticity of markups to income falls to $d\log\mu^{\text{true}}/d\log I \approx 0.110-(0.147)(0.285)-(0.021)(1.265) = 4.2\%$. What is a reasonable proportion of costs to call variable? Using data from a large retail chain with seasonal demand, Kesavan et al. (2014) report that labor hours are 15 percent higher during peak months. If 15 percent of labor and rent costs are variable, the true elasticity of markups to income is 10.0 percent, implying a bias around 1.0pp.

Appendix E Evidence on Search Behavior

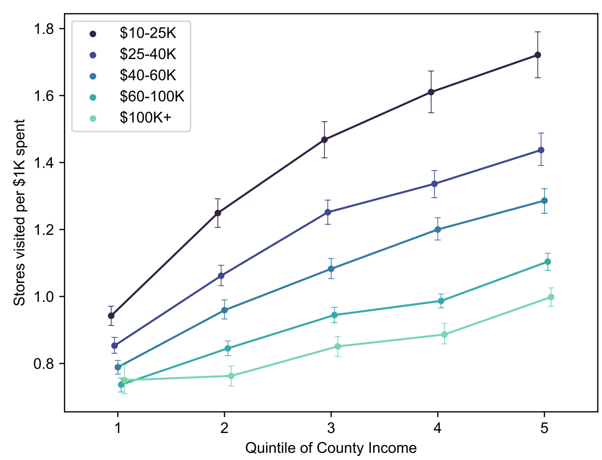
This appendix shows that search intensity in the data is consistent with two predictions of the model: (1) search intensity is decreasing in household income (Lemma 1), and (2) conditional on income, search intensity is increasing in high-income areas (Lemma 5).

We use two measures of shopping behavior from Kaplan and Menzio (2015): the number of shopping trips a household makes and the number of unique stores it visits. As discussed by Aguiar and Hurst (2007a) and Pytka (2018), among many others, measures of search time reflect both search intensity and the size of households' consumption baskets, and so to isolate search intensity, we normalize both measures by dollar spent. Of course, using expenditures to control for basket size risks confounding our results with differences in markups paid, so Table E1 repeats the analysis normalizing by number of transactions, number of unique UPCs purchased, and number of unique brands purchased instead.

Figure E1: Search intensity by income group and county income.



(a) Shopping trips per $1K expenditures.



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

Figure E1 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 and split by the income in each household's county. Two patterns emerge, consistent with both predictions of the model. First, across all county quintiles, high-income households exert less search intensity (Lemma 1). Second, conditional on income, households exert greater search intensity in high-income counties, consistent with strategic substitutes in search (Lemma 5). Table E1 tests these relationships formally and shows that these relationships are robust to using various measures of search intensity and to additional controls (e.g., state fixed effects and the density of grocery establishments in each county).

Table E1: Effect of own income and county income on search intensity.

Dependent variable	Log Own Income		Log County Income	
(Measure of search intensity)	Coefficient	SE	Coefficient	SE
OLS estimates:				
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)
Instrumenting for household income:				
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:

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

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%.

Appendix F Comparison to Other Markup Measures

F.1 Production Function Estimation

De Loecker et al. (2020) estimate markups for public firms in Compustat and find that average markups rose dramatically from 1980 to 2016. In this section, we show that markups of public firms estimated by De Loecker et al. (2020) are also positively associated with buyer income, using data on the income firms' customers from Baker et al. (2023).

Baker et al. (2023) estimate the distribution of buyers' incomes for firms using data on debit and credit card spending by two million individuals from 2010–2015. We calculate the average buyer income for each firm in each year as the expenditure-weighted average over the distribution of buyers' incomes in their data. Merging these measures with markups estimated by De Loecker et al. (2020) yields a sample of 378 firms, including 192 retail firms (NAICS 44–45). We also merge data on sales concentration in NAICS-6 industries from the 2012 Economic Census, available for 152 of the 192 retail firms.

Table F1 reports estimates of the elasticity of firms' markups to average buyer income, including year by NAICS-4 fixed effects. Estimates of the elasticity of markups to buyer income range from 23–44 percent, whether in the full sample or limited to the sample of retail firms. The link between markups and buyer income is robust to including measures of sales concentration in NAICS-6 industries from the 2012 Economic Census.

F.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. In standard models of vertical supply chains (e.g., Tirole 1988 Ch. 3), a reduction in the elasticity of consumer demand increases markups along the entire producer chain, and Wu (2022) shows that similar intuitions can hold in a general production network.

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

We construct a sample of matched upstream–downstream firm pairs using data from Compustat Customer Segments (previously used e.g., by Cohen and Frazzini 2008), which compiles firm disclosures about customers that comprise over 10 percent of sales. We identify cases where the downstream customer is a firm (rather than geography or industry group) and construct a crosswalk from downstream firm names in the Compustat Customer Segments data to firm names used by Baker et al. (2023). Finally, we merge these

Table F1: Impact of buyer income and concentration on De Loecker et al. (2020) markups.

Log Production Function Markup	All (1)	Retail Firms (NAICS 44–45) (2)	(3)	(4)	(5)	(6)
Log Avg. Buyer Income	0.229**	0.358**	0.439**	0.444**	0.443**	0.444**
	(0.085)	(0.067)	(0.094)	(0.094)	(0.094)	(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 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. (2023). The sales shares of top firms in each NAICS-6 industry are from the 2012 Economic Census. Regressions weighted by consumer spending from Baker et al. (2023), and standard errors are two-way clustered by firm and year. * indicates significance at 10%, ** at 5%.

Table F2: Relationship between buyer income at downstream firms with markups of upstream suppliers.

Markup at Upstream Firm	(1)	(2)	(3)	(4)
Log Avg. Buyer Income of Downstream Firm	0.103	0.078**	0.085**	0.076**
	(0.119)	(0.033)	(0.031)	(0.033)
Year FEs	Yes	Yes	Yes	Yes
Year-Upstream Industry FEs		Yes	Yes	
Year-Downstream Industry FEs			Yes	
Year-Upstream Industry-Downstream Industry FEs				Yes
N	9092	8919	8484	7765
R^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. (2023). Upstream and downstream firms are matched using data from Compustat Customer Segments. * indicates significance at 10%, ** at 5%.

firm pairs with markups for the upstream firms estimated by De Loecker et al. (2020).

Table F2 reports how markups at the upstream firms relate to average buyer income at downstream firms. 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. For 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. Of course, these regressions measure correlations in the cross-section of firms. Further work is required to isolate causal effects of changes in price sensitivity downstream on upstream firms' markups.

F.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, we estimate markups at the retailer-UPC level in a single product module (margarine). We choose margarine since the PromoData has especially high coverage in this category: more than 70 percent of margarine sales in the Homescan panel in 2007 are matched to wholesale costs in PromoData. We find that marginal costs and markups recovered from demand estimation are positively correlated with our wholesale costs and retail markups.

Scope and caveats. 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. For this exercise, we assume that manufacturers and retailers maximize joint surplus and divvy that surplus using (unmodeled) two-part tariffs or rebate arrangements. This means that retail markups should form part of the overall markup recovered from demand estimation.

We define products as a retailer-UPC pair and estimate markups and marginal costs separately for each geographical market and each month. As a result, while we do not explicitly model multi-category effects and consumer benefits of one-stop shopping (see e.g., Thomassen et al. 2017), differences in the benefits consumers experience from shopping at different retail chains should be captured in our estimation by quality shifters across retailers. Our approach of estimating markups within one product category at a time is consistent with a broad range of previous studies in industrial organization (e.g., Nevo 2001; Villas-Boas 2007; Nakamura and Zerom 2010) as well as recent work studying the evolution of markups in scanner data (e.g., Brand 2021; Döpper et al. 2021).

Data construction. For data on prices and market shares, we use NielsenIQ Retail Scanner data from 2006–2009, which includes weekly prices and sales of margarine at each store in NielsenIQ's Retail Scanner program. We define geographic markets using designated market areas (DMAs) provided by Nielsen. For parsimony, we consider the 40 DMAs with the largest margarine volume sold from 2006–2009: these markets account for over half of all margarine sold in the NielsenIQ Scanner data over this period. For consumer demographics, we use demographics of NielsenIQ Homescan panelists in each DMA from 2006 to 2009, weighted by Nielsen projection weights.

We define a product as a unique retailer-UPC pair. As noted by Broda et al. (2009), retail chains offer different amenities, so consumers may see the same UPC offered at different retail chains as distinct products. To avoid overfitting to retailers and products with very small sales, we limit our analysis to the largest forty retailers by sales in the sample and the top 150 UPCs by total sales, and combine the remaining products with the outside good. These top UPCs account for over 80 percent of the total volume of margarine sold in the DMAs in the sample. The final sample includes 338,726 observations for prices and volume across 40 DMAs in 48 months.

We construct ownership matrices using brand identifiers from NielsenIQ. The UPCs in our sample belong to 20 distinct brands. Using GS1 parent companies to construct ownership matrices does not qualitatively change the results we present here.

Since prices are endogenous, estimating the demand system requires instruments that are orthogonal to movements in demand. Following Villas-Boas (2007), we use input prices—i.e., monthly prices of soybeans, corn, oil-producing crops, and milk from the USDA National Agricultural Statistics Service—as instruments for prices.

Finally, to define market sizes, we use the NielsenIQ Homescan data to calculate average margarine consumption per household. We take the market size in each DMA in each month to be 0.95 pounds of margarine per household per month, times the number of households in the DMA. Constructed this way, the market share of the outside option displays considerable variation across time and space, ranging between 10 and 90 percent.

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

$$u_{ijmt} = \alpha_i^0 + \alpha_i^p p_{jmt} + \kappa_j + \delta_{jmt} + \varepsilon_{ijmt}, \tag{19}$$

Table F3: Demand system estimates.

	Logit OLS (1)	Logit IV (2)	Random coefficients (3)
Price	-0.622	-1.063	-1.330
	(0.023)	(0.053)	(0.159)
Π_{z0}			-10.036
			(0.131)
Π_{zp}			1.324
			(0.274)
Median price elasticity	1.12	1.91	3.77
Median estimated markup	2.01	1.81	1.41
Product (retailer-UPC) FEs	Yes	Yes	Yes
DMA FEs	Yes	Yes	Yes
N	338726	338726	338726

Note: Estimated using `pyblp` (Conlon and Gortmaker 2020). Standard errors clustered by product. Columns 1–2 are results from the logit specification (20), and column 3 reports results from the random coefficients specification (19).

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, κ_j reflects utility from time-invariant product (retailer-UPC) characteristics, δ_{jmt} is an unobserved demand shifter that varies across across products, markets, and time periods, and ε_{ijmt} is an idiosyncratic draw from a Gumbel distribution. Households in market m and time t purchase one unit of the product that gives them the highest utility (which may be the outside option).

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

$$\alpha_i^0 = \bar{\alpha}^0 + \Pi_{z0} z_i + \Sigma_0 \nu_i^0,$$
$$\alpha_i^p = \bar{\alpha}^p + \Pi_{zp} z_i + \Sigma_p \nu_i^p,$$

where $\bar{\alpha}^0$ and $\bar{\alpha}^p$ are means across households, z_i is log of household i's income, and Π_{z0} and Π_{zp} are interactions that allow the coefficients to vary systematically with income. The random draws ν_i^0 and ν_i^p, which are drawn from standard normal distributions and scaled by standard deviations Σ_0 and Σ_p, allow more flexibility for the model to match patterns of substitution across products.

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

$$\log s_{jmt} = \alpha^0 - \alpha^p p_{jmt} + \gamma_j + \phi_m + \delta_{jmt}, \tag{20}$$

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

Column 1 of Table F3 reports results from estimating the logit model using OLS. The coefficient on price α^p is negative, yielding a median price elasticity of 1.12. Using soybean, corn, oil-producing crop, and milk prices as instruments in column 2 yields a larger coefficient on price, a median price elasticity of demand of 1.91, and a median markup of 1.81.

Column 3 of Table F3 reports results from estimating the random coefficients model. The estimated coefficient on price is significantly negative and similar in magnitude to the logit specification (column 2). The interaction of the value of the outside option with income, Π_{z0}, is negative, which means that high-income households are estimated to gain less utility from consuming margarine relative to an outside option. The estimated interaction of the price coefficient with income, Π_{zp}, is significantly positive, suggesting that high-income consumers are less price sensitive.[33] Incorporating consumer heterogeneity results in a higher median price elasticity of demand of 3.77.

We calculate the average marginal cost for each UPC in each month from the demand model and merge these monthly cost estimates into the NielsenIQ Homescan data for comparison with the wholesale costs and retail markups in the main text. All markup and unit marginal costs in the merged dataset are winsorized at the 1 percent level.

Panel A of Table F4 show that marginal costs recovered from demand estimation and wholesale costs from PromoData are closely correlated (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 ($\rho \approx 0.6$).

Panel B of Table F4 shows that marginal costs of margarine recovered from demand estimation are about 30 cents per pound lower than the base wholesale costs and 10 cents per pound lower than deal wholesale costs in the PromoData. These marginal costs move slightly more than one-for-one with wholesale costs. Within-UPC variation in wholesale costs and retail markups (in panel C of Table F4) also predicts within-UPC variation in the costs and markups recovered from demand estimation.

Our estimates support our assumption of limited heterogeneity in costs of a product across retailers. We regress our estimates of log marginal cost across all retailer-UPC-

[33]Our procedure also yields estimates of Σ_0 and Σ_p, which allow for variation in the utility of the outside option and price sensitivity across households. Both estimates are not significantly different from zero.

Table F4: Relationship between estimated marginal costs and wholesale costs.

Panel A. Correlation coefficients.	Log unit wholesale cost		Log retail markup	
Using PromoData price:	Base	Deal	Base	Deal
Log unit marginal cost (demand est.)	0.92	0.90		
Log markup (demand est.)			0.54	0.62

Panel B. Marginal costs.	Unit marginal cost ($/lb.) (demand est.)			
	(1)	(2)	(3)	(4)
PromoData unit wholesale cost (base)	1.224**	0.248*		
	(0.091)	(0.139)		
PromoData unit wholesale cost (deal)			1.183**	0.303**
			(0.085)	(0.067)
Constant	-0.338**		-0.092	
	(0.106)		(0.092)	
UPC FEs		Yes		Yes
N	261552	261552	261552	261552
R^2	0.92	0.99	0.87	0.99

Panel C. Markups.	Log markup (demand est.)			
	(1)	(2)	(3)	(4)
Log retail markup (using base wholesale cost)	0.743**	0.970**		
	(0.111)	(0.015)		
Log retail markup (using deal wholesale cost)			0.794**	0.960**
			(0.112)	(0.016)
Constant	0.327**		0.134	
	(0.121)		(0.110)	
UPC FEs		Yes		Yes
N	261552	261552	261552	261552
R^2	0.29	0.90	0.38	0.90

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

month-DMA combinations, winsorized at the 1 percent level, on UPC-month fixed effects. These fixed effects explain 80.9% of the variation in log marginal costs. In our empirical analysis in the main text, we further control for county of purchase and for a set of retailer–product characteristics denoted by W_k. Controlling for DMA code increases the share of variation in log marginal costs explained to 83.8%, and further adding W_k increases it to 84.0%. Extending W_k to include quadratic and cubic terms on retailer size (as we do in the robustness check in Table 1) does not increase the share of variation explained. Finally, we find that location and retailer–product characteristics account for the majority of differences in marginal costs across retailers: including fixed effects for retail chain increases the share of variation explained to 85.8%.

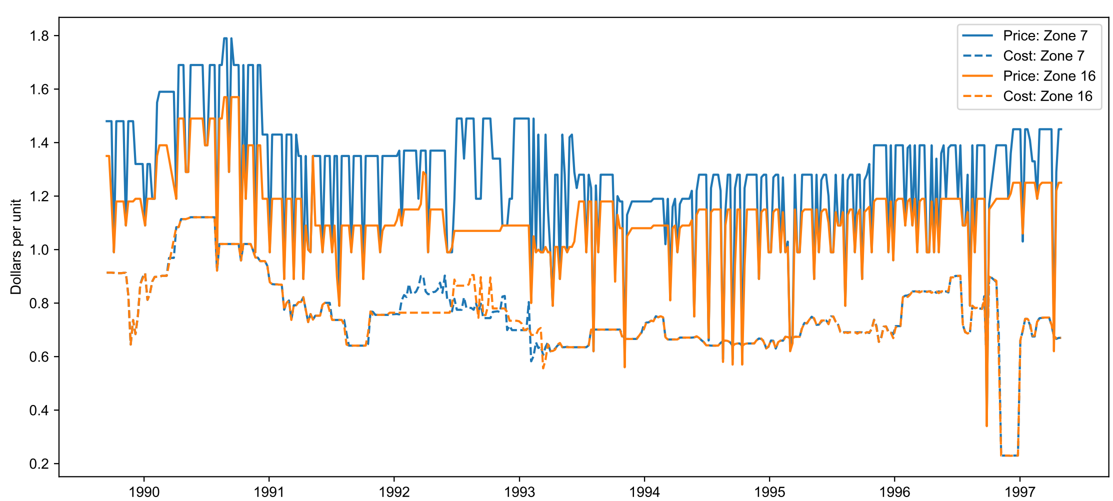
Appendix G Internal Retailer Data

In this appendix, we use data from a retail chain made available by the Kilts Center. These data include both prices and the retailer's internal measures of wholesale costs. We use these data to assess our assumption of cost uniformity across stores and to estimate how accounting for private label products would affect our estimates of the markup gap across income groups.

G.1 Variation in Markups vs. Costs Across Chain Locations

The data include weekly prices and profit margins (and thus wholesale costs) for products in twenty-eight categories at each of the retail chain's stores. These stores are grouped in several different pricing zones: weekly prices across stores within a zone are roughly uniform, but prices set across zones vary. Figure G1 plots the time series of prices and wholesale costs for one example UPC across two different pricing zones. We see that while stores in the two zones charge different prices for the UPC, they record nearly identical wholesale costs for the same UPC. This observation is consistent with our assumption that differences in retail prices across stores primarily reflect differences in markups rather than costs.

Figure G1: Example: Price and cost of a UPC across two pricing zones.



To further test this assumption, we calculate the deviation in the price and wholesale cost of each UPC against the average across all stores in each week. We regress these UPC-week price and cost deviations on demographic attributes around each store (these store-level demographics were created by Market Metrics using US Census data for the

Table G1: Variation in prices and wholesale costs across stores.

	Deviation in Log Price			*Deviation in Log Wholesale Cost*		
	(1)	(2)	(3)	(4)	(5)	(6)
% households value over $150,000	0.0134**			0.0003		
	(0.006)			(0.0002)		
% households value over $200,000		0.0241**			0.0004*	
		(0.007)			(0.0002)	
Log Mean Household Value			0.0080*			0.0002
			(0.004)			(0.0001)
N (millions)	92.7	92.7	92.7	92.7	92.7	92.7
R^2	0.01	0.02	0.01	0.00	0.00	0.00

Note: Standard errors two-way clustered by store and product category. * indicates significance at 10 percent, ** at 5 percent.

Chicago metropolitan area). Table G1 shows results from regressions of price and cost deviations on the percent of households around the store with values over $150,000 and $200,000, as well as the mean household value estimated by Market Metrics.

We find that the degree to which prices across stores covary with store demographics is more than 40 times larger than the degree to which costs vary across stores. Thus, the vast majority of differences in prices for the same UPC across stores is due to markups, not wholesale costs.

Note that the retailer's wholesale costs are not exactly replacement costs (Peltzman (2000) describes the cost accounting used by the retailer), and that they may also omit local inputs that are relevant for marginal costs. These are measurement challenges we also confront in our estimation, and using data from a retailer does not solve these issues.

G.2 Private Label Products and Selection

We estimate how the markup gap across income groups would change if we were able to observe wholesale costs for the entire set of Homescan purchases. First, we use the retailer data to estimate the difference in retail markups that retailers charge for private label products compared to third-party products. Second, we use this estimate, as well as the differences in relative prices that households pay for products not matched to the PromoData, to estimate the markup gap on the entire set of Homescan purchases. Compared to our baseline 13.1pp markup gap, accounting for the higher private label retail markups decreases the markup gap to 11.9pp, and accounting for both private label and selection into our sample increases the markup gap slightly to 13.4pp.

Table G2: Retail markup premium for private label products.

Log Retail Markup	(1)	(2)	(3)	(4)
Private Label Indicator	0.190**	0.188**	0.188**	0.187**
	(0.041)	(0.042)	(0.042)	(0.043)
Product Category–Unit FEs	Yes	Yes	Yes	Yes
Year FEs		Yes	Yes	Yes
Store FEs			Yes	Yes
Controls for log unit price, log package size		Yes	Yes	
Controls for log unit price, log package size interacted with product category				Yes
N (millions)	3.82	3.76	3.76	3.76
R^2	0.37	0.39	0.39	0.41

Note: An observation is a UPC at a store in a given year (aggregated across all weeks). Standard errors are clustered by product category. ** indicates significance at 5 percent.

Estimating the private label markup premium. The PromoData does not include wholesale costs for private label products. This could be a problem, since Barsky et al. (2003) report that private label products tend to have higher retail markups, and Table G3 shows that low-income households spend more on private label products in the Homescan data.

We estimate the private label markup premium by comparing retail markups for private label products at the retail chain to other products in the same product category. Table G2 column 1 shows that our most simple specification, which compares private label products to products in the same category and in the same units (e.g., ounces or liters), indicates a 19 percent premium on retail markups for private label products. Controlling for store, year, the product's unit price, and the product's package size yields similar estimates of an 18–19 percent private label premium.

We estimate how private label products and selection affect the markup gap. Our first adjustment only corrects for the share of expenditures on private label products by each income group,

$$\mu_i^{\text{PL}} = \underbrace{(1 - \phi_i^{\text{PL}})\mu_i}_{\text{Non private label}} + \underbrace{\phi_i^{\text{PL}}\mu_i(1 + P)}_{\text{Private label}}, \tag{21}$$

where ϕ_i^{PL} is income group i's share of expenditures on private label products, μ_i is income group i's average markup in our merged sample (we use the markup differences from Figure 2b, which control for county and retailer–product characteristics, plus a 1.29 markup for the lowest-income group), and $P = 0.19$ is the retail markup premium for private label products that we estimate in the retailer's data.

Table G3: Estimating the effects of private label products and selection on markup gap.

Household Income	Private label share	PromoData match share	Relative unit prices Matched	Relative unit prices Unmatched	Markup gap (pp) Baseline	Markup gap (pp) PL	Markup gap (pp) PL + Selection
<$20K	0.183	0.381	-0.035	-0.059	0.000	0.000	0.000
$20K	0.176	0.379	-0.006	-0.032	1.105	0.957	1.039
$30K	0.169	0.377	0.006	-0.012	1.393	1.084	1.574
$40K	0.161	0.378	0.028	0.013	2.080	1.606	2.407
$50K	0.156	0.373	0.046	0.037	3.054	2.467	3.334
$60K	0.153	0.369	0.071	0.057	3.993	3.360	4.240
$70K	0.143	0.366	0.101	0.098	5.264	4.412	5.682
$100K	0.136	0.353	0.151	0.149	7.932	6.964	8.016
$125K	0.133	0.348	0.169	0.176	8.790	7.785	9.097
$150K	0.124	0.345	0.188	0.207	9.713	8.488	10.138
$200K	0.123	0.324	0.250	0.274	13.107	11.926	13.379

Note: Baseline estimates for the markup gap are from Figure 2b. PL uses the adjustment in (21) and PL + Selection uses the adjustment in (22), using a private label premium of $P = 0.19$ (see Table G2).

These estimates do not account for the fact that products that are not matched to the PromoData have a larger covariance of relative unit prices with income than products in our merged sample. To account for both selection and the private label premium, we regress markups paid by households on relative unit prices of matched products and use the fitted coefficients to predict markups for unmatched products, which we denote $\widehat{\mu}_i$ (relative unit prices of matched and unmatched products are shown in Table G3). We use the extrapolated markups $\widehat{\mu}_i$ and the private label premium $P = 0.19$ to estimate the markup gap accounting for both private label and selection,

$$\mu_i^{\text{PL+Selection}} = \underbrace{\phi_i^{\text{PromoData}} \mu_i}_{\text{Matched sample}} + \underbrace{\phi_i^{\text{PL}} \widehat{\mu}_i (1+P)}_{\text{Unmatched, private label}} + \underbrace{(1 - \phi_i^{\text{PromoData}} - \phi_i^{\text{PL}}) \widehat{\mu}_i}_{\text{Unmatched, not private label}}, \tag{22}$$

where $\phi_i^{\text{PromoData}}$ is the share of expenditures matched to the PromoData.

Table G3 shows the resulting estimates for the markup gap. Compared to our baseline estimates of a 13.1pp markup gap, accounting for the private label markup premium reduces the markup gap to 11.9pp, and accounting for both private label and selection increases the markup gap to 13.4pp.