Rising Retail Concentration: Superstar Firms and Household Demand*

Justin H. Leung† Zhonglin Li‡

November 3, 2022

Abstract

This paper documents an increase in household concentration in the US retail sector from 2004-2019. Despite a growing number of stores, households visit fewer stores, do more one-stop shopping, and increasingly shop at different retailers from each other. We find that the increasing availability of superstar retailers, rises in product variety within stores, and increases in households' opportunity cost of time contribute to these trends. We develop a model that can rationalize these results. Our calibrated model shows that household concentration is tightly linked to markups and that its rise led to a 5 percentage point increase in aggregate markups.

JEL Classification Numbers: L10, L81

*We thank Milena Almagro, Marianne Bertrand, Mark Bils, Austan Goolsbee, Erik Hurst, Ryan Kim, Matt Notowidigdo, Canice Prendergast, Esteban Rossi-Hansberg, Matthew Shum, Chad Syverson, Joseph Vavra, Thomas Wollmann, and seminar participants at the University of Chicago, National University of Singapore, Meeting of the Urban Economics Association, ISMS Marketing Science Conference, and Asian Meeting of the Econometric Society for valuable comments and advice. Justin Leung gratefully acknowledges research funding support from the National University of Singapore. Researcher(s)' 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 researcher(s) 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.

†National University of Singapore (e-mail: bizjhl@nus.edu.sg).

‡National University of Singapore (e-mail: zli@nus.edu.sg).

1 Introduction

Many economists and policymakers are expressing concern over the possibility of increasing monopoly power of superstar firms. An intriguing case is the US retail sector, which has witnessed significant changes in recent decades. National market concentration has risen substantially.[1] However, concentration measures at finer levels of geography have only increased slightly (Smith and Ocampo 2021) or decreased substantially (Hsieh and Rossi-Hansberg 2019; Rossi-Hansberg et al. 2021). This is consistent with the expansion of superstar national retail chains into more geographic locations.[2] Given that the standard approach in retail merger enforcement uses geographic market concentration as the relevant measure of competition (Balto 2001), this suggests that the retail sector has not become substantially less competitive.

However, a unique feature of the retail sector is the provision of convenience by facilitating one-stop shopping. Households may concentrate on different retailers in the same region. Therefore, market-level concentration may barely change even as households engage in more one-stop shopping. Given that the intensity of one-stop shopping is related to market power (Thomassen et al. 2017), this implies that market-level concentration may be an insufficient metric for competition.

In this paper, we provide empirical and theoretical evidence that household concentration can serve as a more informative metric for competition. First, we show descriptive evidence on a rise in household concentration and one-stop shopping. Second, we provide empirical evidence to explore the driving forces for these trends. Third, we develop a model to rationalize our results and highlight the implications for market power.

We begin our analysis by using the Nielsen Consumer Panel Dataset to document trends in household shopping behavior. First, we find that households are shopping more in their preferred stores. From 2004 to 2019, the average household retail HHI has increased by 17% from about 0.35 to 0.41 while real expenditure remained stable. This rise holds within different demographic groups. In accordance with these facts, households visited fewer retailers annually. The number of retailers that an average household visited per year dropped by 15% despite a growing number of retail establishments in local markets. These facts imply that households are not constrained to fewer choices of retailers. Second, households are not only concentrating purchases more in some retailers within product departments, but also shopping for products in different departments increasingly in the

[1] From 2004 to 2019, the market share of the top four retail chains has increased from 25% to over 37%. These numbers are calculated using the Nielsen Consumer Panel data. It covers major categories in the retail sector, including drug, grocery, and merchandise stores. Details are in Section 2.

[2] We denote superstar firms as those with the largest market shares.

same retailer. Households also made fewer shopping trips. Correspondingly, households have been spending more and purchasing more varieties of products per trip, pointing to more one-stop shopping.

What is the relationship between household and market concentration? We show that the regional HHI can be decomposed into the average household HHI minus cross-household variance. The variance term reflects heterogeneity across households in their choice of retailers. We find that cross-household variance has increased by a comparable magnitude to that of within-household retail concentration, keeping the local retail concentration relatively unchanged. This signifies that different households are actually concentrating on different retailers over this period.

What are the driving forces behind these facts about household consumption? We attempt to provide causal empirical evidence to address this question. On the supply side, we first investigate the impact of the increasing availability of superstar big-box retailers such as supercenters and club stores. We utilize an event-study approach to estimate the impact of the entry of these superstar retailers. We find that an additional supercenter raises household HHI by about 1 percentage point and increases varieties per trip, while an additional club store actually lowers household HHI by about 1 percentage point and decreases varieties per trip. Using these estimates, we calculate a back-of-the-envelope estimate (BOTE) for how much these entries explain the rise in household HHI. We find that supercenter entries can explain about 17% of the rise in household HHI within regions for which supercenter entries take place, but many regions did not experience a supercenter entry, such that supercenter entries only explain about 2% of the rise in household HHI across the US.

To offer potential explanations for why supercenters and club stores exhibit different effects, we document that households buy far more varieties than any other channel type, including club stores, and pay relatively low prices. This suggests that superstar retailers with larger assortments may allow households to benefit from demand-side economies of scope and engage in one-stop shopping. They may also achieve lower prices by benefiting from economies of scale to attract more customers. These differences are further explored in Leung and Li (2022).

We directly investigate these hypotheses by estimating the effect of changes in product variety and prices within existing stores. We follow the recent literature on uniform pricing and assortment similarity within retail chains (DellaVigna and Gentzkow 2019) and utilize an instrumental variable (IV) based on the variety and pricing of products in other stores of the same retail chain. We estimate that as stores increase their variety, households in the region tend to increase their HHI and varieties per trip. When stores charge lower prices relative to its competitors, households in the region also increase their HHI. Our BOTEs imply that increases in variety over this sample period due to national chain-level changes can explain up to 16% of the increase in household HHI, while changes in prices did not substantially account for the rise in household HHI, since superstar retailers were not charging increasingly lower prices. This provides additional support to the hypothesis that demand-side economies of scope and supply-side economies of scale can both increase retail concentration.

After showing that supply-side extensive and intensive-margin decisions by retailers can explain a portion of the rise of household HHI and one-stop shopping, we now turn to the demand side by studying the impact of changes in time costs among households. We first show using an event study that an episode of non-employment, which should lower time costs, decreases household HHI and raises the number of shopping trips. Next, following the literature on opportunity costs of time, we use variation in household characteristics over time and proxies of labor demand shocks across demographic groups as IVs for the number of shopping trips. We estimate that a decrease in the number of shopping trips raises household HHI, cross-household variance, and varieties per trip. Our BOTEs imply that the decrease in shopping trips has the potential to explain the entire rise in household HHI.

What then are the forces that drive the decrease in shopping trips? We document that the opportunity cost of time, as proxied by average hourly wages and unemployment rate, has risen over the sample period. Our empirical results show that age and education groups that experienced a larger change in the opportunity cost of time decrease the number of shopping trips more. This mechanism can explain up to 14% of the rise in household HHI.[3]

Do these facts reflect a rise in the market power of superstar chains and thus a decline in consumer welfare? Existing literature has generally found evidence that the supply-side developments we document that increase retail concentration tend to raise consumer welfare through economies of scale and scope.[4] However, the demand-side developments are more ambiguous, as a rise in one-stop shopping could be a result of demand-side changes that either increase or decrease consumer welfare.

Therefore, we develop a model that can rationalize our results. We model the demand-side as a two-layer nested CES utility function following Hottman et al. (2016) among others. The first nest contains firms and the second nest contains products, which enables consumers to choose from any number of multi-product firms. We then introduce a cost to visiting each firm following Bronnenberg (2015) and allow consumers to derive utility from leisure. Based

[3]Given the difficulty of finding exogenous shifters of the opportunity cost of time in the literature, we would like to acknowledge that while we extend on the literature by exploiting panel variation at the household level and using an event-study approach, these results are suggestive evidence and may only be interpreted as plausibly causal.

[4]For example, see Atkin et al. (2017) and Leung and Li (2022). We further elaborate on these papers below.

on our model, supply-side changes that increase household retail concentration, such as the increasing prevalence of supercenters and increases in product variety and lower prices by firms, will increase consumer welfare. Demand-side changes that increase the opportunity cost of time and increase household concentration may either increase or decrease welfare, depending on the source of these changes. While increases in the cost of travel will decrease welfare, wage increases will increase welfare. By introducing heterogeneity in shopping costs, our model can also account for a rise in cross-household variance.

We also show that household concentration, rather than market concentration, is tightly linked to markups under oligopolistic competition. Our calibrated model shows that the aggregate markup increased by over 5 percentage points in our sample period. Therefore, these supply- and demand-side developments that increase household concentration and welfare can also increase markups, offsetting some of these welfare gains.

Our paper contributes to several strands of literature. First, our paper contributes to a rich literature on concentration and market power summarized in Berry et al. (2019) and Syverson (2019). They highlight that the theoretical relationship between market concentration and average market power is ambiguous.[5] The literature has documented rising market concentration since 2000 or earlier. One strand points to changes in technology and economies of scale for firms as the dominant explanation (e.g. Autor et al. 2020, Benkard et al. 2021, Rossi-Hansberg et al. 2021, Ganapati 2021, and Kwon et al. 2021). Another strand mainly attributes these trends to firms exercising their rising market power (e.g. Covarrubias et al. 2020, De Loecker et al. 2020, and Grullon et al. 2019). Berry et al. (2019) and Syverson (2019) call for a surge in industry-level research to characterize heterogeneity more fully both across and within markets, suggesting that sources of these patterns may be multi-causal, all with potential implications for market power in possibly different directions. Our paper attempts to explore these multi-causal sources in the US retail sector using micro-data. We provide empirical and theoretical evidence that household concentration can serve as a novel and informative metric for market power.

Second, our paper contributes to a literature focusing on the evolution of the US retail sector in recent decades. Complementing Hsieh and Rossi-Hansberg (2019), who highlight a new industrial revolution in services, retail, and wholesale due to the availability of a new set of fixed-cost technologies that lower marginal costs in all markets, we focus on store entry, variety, and pricing as supply-side mechanisms that can increase both consumer welfare and markups. We also complement Neiman and Vavra (2021), who find similar patterns of increasing household concentration and heterogeneity in product markets. They

[5]Many empirical studies find patterns of simultaneous concentration and productivity growth, and Syverson (2019) argues that the case for large and general increases in market power is not yet dispositive.

conclude that increasing product variety drives these trends. By contrast, we document patterns in retail markets and highlight increases in one-stop shopping. Our results are also consistent with Döpper et al. (2021), who find that consumers are becoming less price sensitive over time, which drives up markups in product markets. Our results show that the rise of household retail concentration can contribute to decreasing price elasticities in retail markets, which potentially relates to rising markups in product markets through double marginalization. Another related paper is Smith and Ocampo (2021), who document trends similar to ours in national and local retail concentration using Census data. In contrast, we are able to explore deeper by studying household concentration with consumer scanner data.

Third, our paper adds to a literature on the impact of big-box stores, as summarized in Carden and Courtemanche (2016) and Ellickson (2016). Hwang and Park (2016) document the impact of Walmart supercenter conversion, as opposed to entry, on varieties per trip, similarly finding that consumers increase one-stop shopping. Atkin et al. (2017) provide reduced-form evidence on the impact of Walmart entry in Mexico and estimate large welfare gains for households. Leung and Li (2022) similarly provide reduced-form evidence on the impact of big-box store entry in the US and quantify various sources of welfare gains using a different model to show product variety is a key differentiating factor between supercenters and club stores. In contrast, this paper focuses on how big-box store entry by superstar firms, along with within-store changes in product variety and pricing by national retail chains, contribute to the rise in retail concentration, which offsets some of the welfare gains by raising markups.[6]

Fourth, our paper contributes to a large literature on time use in economics and marketing. In economics, a wide variety of papers focus on the substitution between time and market goods, highlighting how households trade off shopping time for lower prices through increased shopping effort over the lifecycle and business cycles (e.g. Aguiar and Hurst 2007; Aguiar et al. 2013; Nevo and Wong 2019). Coibion et al. (2021) also document a drop in shopping frequency and highlight the implications for measurement of consumption inequality.[7] In marketing, Bronnenberg (2018) summarizes the literature on how structural changes in consumers' time allocation impact retail strategy, and conversely, how retail innovations that make purchasing and home production more convenient impact purchasing

[6] Basker et al. (2012) show that general merchandisers that added the most stores also made the biggest increases to their product offerings, and explain these facts with a stylized model in which a retailer's scale economies interact with consumer gains from one-stop shopping to generate a complementarity between a retailer's scale and scope. We show that this interaction between retailers and consumers can be used to explain a different set of trends.

[7] Our results are also consistent with the theoretical mechanism proposed by Kaplan and Menzio (2016), who argue that market power goes up when unemployment falls, benefiting firms and leading to self-fulfilling unemployment fluctuations.

habits and time use of consumers.[8] We add causal evidence using panel variation and micro-data on how increases in opportunity cost of time raise demand for one-stop shopping and affects retail concentration. A closely related paper to ours is Bronnenberg et al. (2020), who study how the availability of additional time shifts a households' shopping bundle towards more time-intensive goods in the Netherlands. In contrast, we study how retail strategy and household production interacts to affect retailer choice and increase market concentration in the US.

2 Data

2.1 Nielsen Consumer Panel

The Nielsen Consumer Panel Dataset (henceforth HMS) represents a longitudinal panel of approximately 40,000 to 60,000 US households from 2004 to 2019 who continually provide information to Nielsen about their households and what products they buy, as well as when and where they make purchases.[9] Panelists use in-home scanners to record all their purchases, from any outlet, intended for personal, in-home use. Products include all Nielsen-tracked categories of food and non-food items, across all retail outlets in the US. Nielsen samples all states and major markets. Panelists are geographically dispersed and demographically balanced. Each panelist is assigned a projection factor, which enables purchases to be projectable to the entire US.

Panelists report the products they purchase in each shopping trip. For each product as defined by its universal product code (UPC), we know the quantity purchased and total price paid for all units. Over 5 million products are further classified into about 1100 product modules, 125 product groups, and 10 product departments, which allows us to calculate varieties at various levels. A de-anonymized retail chain identifier is specified for each trip so that we are able to calculate the market share of each retail chain. We also observe where the household resides at various geographic levels from the Nielsen Scantrack market level (Nielsen classifies regions into around 50 market areas) down to the level of county and 5-digit zip code.

[8]Early work includes Messinger and Narasimhan (1997), who develop a model of retail formats based on consumers' increased demand for one-stop shopping and estimate it using time-series variation in aggregate data from 1961-1986. They argue that growing demand for time-saving convenience drives increasing assortment.

[9]The data are available through a partnership between NielsenIQ and the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business. Information on access to the the consumer panel data as well as the retail scanner data described below is available at http://research.chicagobooth.edu/nielsen/.

2.2 Nielsen Retail Scanner

The Nielsen Retail Scanner Dataset (henceforth RMS) consists of weekly pricing, volume, and store merchandising conditions generated by participating retail store point-of-sale systems across the US from 2006 to 2019. Data are included from approximately 30,000-50,000 participating stores and include store types such as drug, grocery, and mass merchandise stores, covering around 53-55% of national sales in food and drug stores and 32% of national sales in mass merchandise stores. The finest location of each store is given at the county level. We use this data to supplement our main analysis using HMS whenever needed, since RMS contains richer information at the store-level, recording every UPC that had non-zero weekly sales in each covered store.

2.3 Store Locations

We obtain the store locations and opening dates of several superstar retail chains from 2004-2013 using data from Arcidiacono et al. (2020) and Coibion et al. (2021).[10] These include Walmart supercenters and three club chains: Costco, Sam's Club, and BJ's. This allows us to conduct event studies to study the impact of superstar big-box retailers, which we describe in detail in Section 4.1.

We also acquire the Nielsen TDLinx data from 2004-2019 to obtain monthly-level store counts for each retail chain at the market level.

3 Descriptive Evidence

We document several motivating facts using the HMS. First, we calculate the HHI at four decreasing levels of aggregation: nationally, regionally at the Scantrack market and county level, and at the household level. We show the specific formulas for measuring HHI in Appendix Section A. Figure 1 shows that national HHI has been rising throughout the entire sample period from 2004 to 2019 by over 3 percentage points. We see a similar but more moderate increase at the market level of about 2 percentage points, but at the county level, the trend becomes almost flat with a roughly 1 percentage point increase. However, a trend that shows the largest increase is recovered at the household level, increasing from around 0.35 in 2004 to almost 0.41 in 2019, a 17% increase (6 percentage points). We also plot the distribution of retail concentration over time along with the weighted mean in Appendix Figure G1, as well as the changes by percentile each year in Appendix Figure G2.

[10]We thank the authors for making their data available to us.

While there are no obvious patterns for the market and county level, the household level shows clearly that the changes are driven by the upper percentiles.

To reconcile these facts, we show that changes in HHI at different levels of aggregation can be linked by a decomposition following Radaelli and Zenga (2002) (RZ). We show in Appendix Section A.1 the exact formulas for this decomposition. In short, the national HHI can be decomposed into the revenue-weighted average of regional HHIs minus the cross-region variance, which captures how different the market share distributions are across regions, as shown in equation (1).

$$\Delta\text{National HHI} = \Delta\text{Regional HHI} - \Delta\text{Cross-region Variance} \tag{1}$$

A larger variance implies a large difference across regions. Hence, a rising national HHI along with a flat county HHI implies that counties are becoming increasingly similar in their market share distributions even as the weighted average of HHI within each county is roughly flat, as shown in Figure 2a.[11]

How are superstar firms linked to these trends? First, we show how many superstar firms are expanding across the nation and opening stores in more markets using the Nielsen TDLinx data in Appendix Table F1. We separate retail chains into two groups: the top 40 retail chains by revenue rank over the sample period (superstar firms), and other chains outside the top 40. The median superstar chain has increased the number of stores by 11% from 2004-2019, while the median chain outside the top 40 has decreased by 7%. We show in Appendix Figure G4 that using the number of stores in each retail chain and market to calculate the HHI gives a similar pattern to using revenue in the HMS, with region HHI and cross-region variance both increasing to contribute to an increasing aggregate HHI. These facts are consistent with findings in Hsieh and Rossi-Hansberg (2019) and Smith and Ocampo (2021).

Next, we show in Appendix Section A.2 that the concentration trends can be decomposed into contributions by various groups of firms. We illustrate how the two groups of chains contribute to rising concentration in Appendix Figure G5. Although superstar firms drive trends in both aggregate HHI and region HHI, changes in cross-region variance is driven by both groups of firms. One potential reason is that the expansion of chains into more geographic locations can either increase or decrease cross-region variance.[12] As shown in

[11] We show these decompositions for the market level in Appendix Figure G3.

[12] For intuition, consider a chain with one store in one region. The cross-region variance will be very small as the regional market share is zero in almost every region except the one in which the chain has a store, and the national market share is also close to zero. As a chain expands into more geographic locations gradually, more and more locations will have non-zero market shares and the national market share will also increase. Hence, cross-region variance will increase. At a certain point, the number of regions without stores will

Appendix Table F1 and F2, there is considerable dispersion in cross-region variance across both superstars and other chains.

The RZ decomposition can be applied at each lower level of aggregation. We further decompose county HHI into household HHI and cross-household variance as in equation (2).

$$\Delta\text{County HHI} = \Delta\text{Household HHI} - \Delta\text{Cross-household Variance} \tag{2}$$

A rising household HHI is consistent with a flat county HHI when households increasingly buy at their preferred retailers while different households increasingly concentrate on different retailers. This leads to an increase in the cross-household variance as shown in Figure 2b. In this paper, we focus on this novel finding at the household level.

We investigate whether these trends are driven by households with certain demographic characteristics. Figure 3 and Appendix Figure G6 plots the changes and levels in household HHI for various household income groups, households living in more urban vs. rural counties, and various age and employment status groups by the gender of the household head. While we find differences in the levels of HHI across different demographic groups, we find that the HHI is increasing in nearly every demographic group albeit at different speeds. For example, households living in more rural areas and younger households experience a larger increase in HHI.[13] We also show in Appendix Section B that our results are robust to controlling for potential changes to the sample of households and that the proportion of consumption that our sample captures relative to the Consumption Expenditure Survey has been stable over time.

Does the increase in national and household HHI imply a decreasing availability of retailers for households? We suggest that this is not the case in Figure 4. While the rise in household HHI is indeed driven by a 15% decrease in the number of retailers visited each period, the real expenditure has barely contracted over the sample period. The number of drug, grocery, and mass merchandise retail establishments per county has also risen fairly substantially over this period. We also find that households are decreasing their frequency of shopping trips, spending less days per week shopping.

We then further decompose household HHI into household-product-category HHI and cross-category variance in Figure 2c, where we define each product category at the

diminish, and cross-region variance may decrease. Each region becomes more similar as the whole nation becomes more saturated with stores from this chain. Therefore, it is likely that as a chain expands across the nation region by region, cross-region variance will first increase and then decrease, unless stores are opened separately in each market in a relatively uniform fashion. We find that in our data, it is indeed the case that some superstar firms first experience a rise in cross-region variance and then a fall in cross-region variance.

[13] We further investigate classifications by other demographic groups in Appendix Figure G7.

product-department level as in equation (3).

$$\Delta\text{Household HHI} = \Delta\text{Household-category HHI} - \Delta\text{Cross-category Variance} \tag{3}$$

We find that the household-category HHI is increasing while the cross-category variance is decreasing, implying that households are increasingly buying different product categories at the same retailer.[14] We show in Figure 5 that households are indeed spending more per trip. They do this partly by increasing the number of varieties per trip, whether measured by the number of UPCs, brands, product modules, or product groups.

These changes in the number of trips and retailers visited by households can be linked as shown in Appendix Figure G12. The number of trips can be decomposed as the number of retailers multiplied by the number of trips per retailer. We show that the drop in the number of trips is driven entirely in the drop in retailers visited as opposed to a drop in trips per retailer. Real expenditure can be decomposed as either the number of retailers multiplied by the real expenditure per retailer or the number of trips multiplied by the real expenditure per trip. The real expenditure per retailer and trip has increased, consistent with our previous findings.

Why are households increasingly shopping in their preferred retailers but concentrating on different retailers? Why are they decreasing their shopping trips while increasing their expenditure and varieties per trip even as the number of retail establishments has increased? We turn to providing plausibly causal evidence to explain these facts in the next section.

4 Reduced-Form Evidence

In this section, we present reduced-form evidence for a series of potential explanations for the trends we observe in the previous section. In each subsection, we list out both the empirical strategy as well as the results. On the supply side, we first analyze the effect of the entry of several big-box superstar retailers in Section 4.1. We then investigate the impact of changes in product variety and prices by retail stores in Section 4.2. On the demand side, we study the impact of changes in time costs among households in Section 4.3. We discuss other hypotheses in Section 4.4.

[14]We show that these results are not driven by compositional changes in our dataset by also using a Dynamic Olley-Pakes decomposition (Melitz and Polanec 2015) on top of the RZ decomposition in Appendix Table F3, with the details shown in Appendix Section A.1.

4.1 Entry of Superstar Big-box Retailers

4.1.1 Empirical Strategy

To study the impact of the increasing availability of superstar big-box retailers such as supercenters and club stores, we utilize an event-study approach to estimate the impact of the entry of these superstar retailers. Our baseline independent variable measures the number of stores for each chain within the 5-digit zip code of each household. We also calculate alternative distance measures. As shown in equation (4), we then regress our outcome of interest for household i in quarterly period t, for example the household retail HHI, on the number of stores, and add household fixed effects to control for fixed household characteristics, as well as period fixed effects to control for national time trends.

$$Y_{it} = \beta \times Num_{it} + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{4}$$

If a store enters in periods when unobservable local household characteristics change, or households anticipate these openings by changing patterns in significant ways, then this would be a threat to our identification. A priori, we believe that it is difficult for households or stores to exactly time sharp changes in unobservables with store entry. To further alleviate these concerns, we estimate the trends before and after the entry event by adding leads and lags of the independent variable Num_{it}. If the pre-trends are parallel, we argue that this gives additional evidence to suggest that stores or households have limited ability to anticipate the precise timing of the entry.

4.1.2 Results

We show the results of estimating equation (4) in Table 1. We estimate the effect of an additional supercenter or club store respectively in a household's 5-digit zip code. We also separately estimate the effect for only zip codes with one entry event or all zip codes. We find that an additional supercenter raises household HHI by 0.008 while an additional club store lowers household HHI by 0.012. These results are statistically significant at the 1% level.[15]

We calculate a back-of-the-envelope estimate (BOTE) of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample

[15] Note that only about 1% of households live in a zip code that ever had two or more supercenters or club stores, such that these coefficients are almost all identified by households that go from zero to one store.

period, then dividing this number by the total change in household HHI in the sample period. Leung and Li (2022) and Hortaçsu and Syverson (2015) both document the dramatic rise of supercenters and club stores over this period. We find that in zip codes with at least one entry, the rise in the number of supercenters explains about 17% of the rise in household HHI, while the rise in club stores decreases HHI, explaining about -8% of the rise in household HHI. Since only 6% and 14% of households experienced a supercenter and club store entry respectively, the BOTE is much smaller at about 2% and -1% respectively for all regions.

In Figure 6, we show that pre-trends are parallel around store entries roughly two years before the event, while the effects are dynamic and continue to rise for an extended period after the event. This is consistent with both households taking time to learn about the presence of new stores and adjusting their purchasing habits. We also show that our results are robust to using alternative measures of distance in Appendix Table F4 such as distance to the nearest store or the number of stores within a certain mile radius. The fact that the effect dissipates as the distance from each household rises increases our confidence in a causal interpretation of our estimates. Our results are also similar when using an estimator robust to heterogeneous treatment effects from de Chaisemartin and D'Haultfœuille (2020) (hereafter DCDH), as shown in Appendix Figure G13.

To provide suggestive evidence of why supercenters generate a different effect from club stores, we show how supercenters and club stores differ in two characteristics that typically define a retailer: product variety and prices.

In Figure 7, we show the average number of UPCs and product modules per household-quarter for supercenters, club stores, and other channel types over the sample period. Households buy far more varieties in supercenters than any other channel type, whereas club stores sell fewer UPCs and product modules compared with grocery stores and supercenters, with the number of varieties close to other channel types such as discount stores and dollar stores, but higher than drug stores and other miscellaneous channel types. This would be consistent with the hypothesis that superstar retailers with larger assortments allow households to benefit from demand-side economies of scope and engage in more one-stop shopping, increasing their household HHI. We directly investigate the effect of variety in Section 4.2.

In Figure 8, we calculate the relative price index (RPI) of each retailer or channel type following Aguiar and Hurst (2007) (hereafter AH). To construct a retailer RPI, we calculate the ratio between total expenditure for each good and the counterfactual expenditure of each good at its average price in the reference region. We then take the weighted average across goods and counties to calculate a national RPI for each retailer that uses national averages as reference prices. We find that supercenters consistently offer lower prices than its competitors nationally, although their price advantage has been decreasing. While club stores generally have an RPI below one, they tend to have higher RPIs than supercenters over the sample period, with the exception of Club 2 offering lower prices in the last periods, which our entry data does not capture. This would be consistent with the hypothesis that superstar retailers with lower prices may attract more households to benefit from supply-side economies of scale, increasing their household HHI. However, ex-ante it is difficult to predict how prices mediate the effect of big-box store entry on household HHI, since household HHI may be concentrated before store entry precisely because low-price alternatives are not available. We directly investigate the effect of prices in Section 4.2.

We investigate the effect of entry on other measures of concentration in Appendix Table F5. We find that the entry of supercenters decreases the number of retailers visited by households while club stores have the opposite effect. The rise in household HHI due to supercenter entry is driven both by a rise in household-category HHI and a drop in cross-category variance, while the drop in household HHI due to club entry is driven mostly by a drop in household-category HHI. Cross-household variance decreases but the change is statistically insignificant.

We estimate the effect of entry on the number of trips per quarter and varieties per trip for households in Appendix Table F6. Supercenters actually do not decrease the number of trips by a significant amount while clubs increase the number of trips by a significant amount. Nonetheless, supercenters do increase the number of varieties per trip, in particular the number of departments per trip, while club stores have the opposite effect. This is once again consistent with larger assortments in supercenters relative to club stores.

How do these effects vary across households? We approach this question in two ways. First, we use tools from Chernozhukov et al. (2018) to explore heterogeneous treatment effects across households. To do this, we add interaction terms to equation 4 using variables such as household income, size, age, and region, to estimate sorted effects. We then analyze whether characteristics are different across the most affected and least affected households. We plot the sorted effects in Appendix Figure G14. While the effects are indeed heterogeneous, we find that most of the differences in household characteristics between the most and least affected groups are not statistically significant. The point estimates do suggest that lower-income, younger, single households living in more rural areas are more likely to increase concentration when facing a supercenter entry. However, even for these households, BOTEs suggest that supercenter entry does not fully explain the rise in household HHI. Second, we use a residualized quantile regression following Borgen et al. (2021). We find that the effect is indeed largest at the upper percentiles as shown in Appendix Figure G15, consistent with our finding that the rise in household HHI is driven by households in the upper percentiles of the distribution. The BOTEs implied by these coefficients are similar to those using equation (4).

Overall, we find that supercenters do increase household HHI but they do not explain the entire increase over the sample period. This is true both because supercenter entries can explain only about 17% of the rise in household HHI within regions for which supercenter entries take place, and because many regions did not experience a supercenter entry. On the other hand, club stores actually work in the opposite direction. This suggests that supply-side changes in the retail landscape as measured by entry of superstar big-box retailers contribute only partly to the rise in household HHI.

4.2 Variety and Prices

4.2.1 Empirical Strategy

To estimate the effect of changes in product variety and prices by retailers on households, we follow the recent literature on uniform pricing and assortment similarity within retail chains (DellaVigna and Gentzkow 2019) and utilize an instrumental variable (IV) that is based on the variety and pricing of products of other stores in a given retail chain. This IV strategy relates to the one in Hausman and Bresnahan (2008) and has been employed recently in DellaVigna and Gentzkow (2019) and Allcott et al. (2019) among others.

Specifically, our estimating equation is as follows:

$$Y_{it} = X'_{rt}\gamma + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{5}$$

We regress our outcomes of interest such as household retail HHI on a vector of variety and prices for each region-period rt. This vector includes two variables which are logged region-level revenue-weighted averages: (1) product variety, as measured by the number of UPCs or product modules per store, and (2) the price index for each store. Each variable x_{rt} is constructed as the revenue-weighted average for each store s in region r:[16]

$$x_{rt} = \sum_{s \in r} w_{st} x_{st} \tag{6}$$

To isolate variation coming from supply-side changes that are plausibly exogenous to unobservables that affect local household outcomes, we construct an IV z_{rt} excluding all

[16] In the HMS, we only observe each store as a retail chain-region pair cr.

stores in region r:

$$z_{rt} = \sum_{c \in r} w_c \frac{\sum_{s \in c} \left(w_{st} x_{st} - \sum_{s' \in r} w_{s't} x_{s't} \right)}{\sum_{s \in c} \left(w_{st} - \sum_{s' \in r} w_{s't} \right)} \tag{7}$$

For each retail chain c, we first construct its revenue-weighted national average of the variable leaving out the region of interest. We then weight it by the revenue share it earns in that region in the entire sample period w_c, allowing us to hold the weight fixed across time. Therefore, the identification assumption is that retailers price and stock products similarly across their chains, such that national supply shocks to the chain affect local prices and assortment, but are plausibly exogenous to unobservable demand shocks that affect our household outcomes.

4.2.2 Results

We show the results of estimating equation 5 in Table 2. We estimate the effect of a percentage change in variety and prices respectively. We construct these variables using both the HMS and the RMS for comparison, harnessing the strength of each dataset. For the HMS, we construct an RPI as described in Section 4.1.2 since it has broader cross-sectional coverage nationally. We use both region (county) and national reference prices. For the RMS, we construct a store price index following Leung (2021) due to its ability to observe products at higher frequencies in each store. Likewise, for the variety measure, the HMS will be able to capture a broader set of retailers while the RMS is able to capture variety within store more precisely for the set of stores it contains.

We find that increasing variety as well as lower prices both lead to a rise in household HHI. This is true particularly for the number of UPCs. These results are consistent with results in the previous section, where entry of supercenters with both more variety and lower prices increases household HHI. This also implies that demand-side economies of scope and supply-side economies of scale can both increase retail concentration.

Our BOTEs imply that increases in variety over this sample period due to national chain-level changes can explain 3-16% of the increase in household HHI, while changes in prices did not substantially account for the rise in household HHI.[17] This is because as seen in Figure 8, the largest retailers were not offering increasingly lower prices relative to competitors. We show that nationwide, stores are indeed offering more varieties in Figure

[17] To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively.

G16, especially those with higher store revenues. Stores with more variety have a particularly strong effect on household HHI when they increase variety as shown in Appendix Table F9, explaining about 22% of the rise in household HHI.

We also estimate equation 5 for other measures of concentration in Appendix Table F10 and F11. These results remain consistent with Table 2. Increases in variety and lower prices decreases the number of retailers visited. More variety and lower prices increases household-category HHI and decreases cross-category variance, although less so for variance. The effect on cross-household variance is small overall.

We estimate the effect of variety and prices on the number of trips per quarter and varieties per trip for households in Appendix Table F12 and F13. Increasing variety significantly increases the number of varieties per trip, while changing prices has a more mixed and much smaller effect. The effect on the number of trips is mixed across the datasets, but is roughly consistent with a rise in variety contributing to the decrease in the number of trips, as shown by the BOTE in Table F13.

4.3 Time Costs

4.3.1 Empirical Strategy

To estimate the effect of changes in the opportunity cost of time on households, we first motivate our empirical strategy by conducting an event study. We regress household retail HHI or the number of shopping trips on an indicator for whether the female or male household head is not working, focusing on prime-aged household heads from age 25-54. Household and period fixed effects are included. We use both OLS and methods robust to heterogeneous treatment effects following DCDH. Given that changes in employment status may be correlated with changes in unobservables affecting the outcomes of interest, this event study allows us to visually examine whether pre-trends are parallel.

Next, we follow AH and use the number of shopping trips per household in each time period as a proxy for the amount of shopping time spent. Figure 5 in Coibion et al. (2021) offers support for the use of this proxy, since it shows similar trends in shopping time using the American Time Use Survey (ATUS). We then evaluate how plausibly exogenous changes in shopping time affect household retail concentration. Specifically, our estimating equation is as follows:

$$Y_{it} = \beta \times N_{it} + X_{it}'\gamma + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{8}$$

We regress our outcomes of interest such as household retail HHI on the number of shopping trips N_{it} for each household-period it along with a vector of control variables. We then instrument for the number of shopping trips with a series of IVs following AH and Aguiar et al. (2013).

First, we use household characteristics as IVs. By including household and time fixed effects, we can extend beyond using cross-sectional variation as in AH and leverage within-household variation in household characteristics, using household age, size, and employment status as IVs. A caveat is that these IVs themselves may reflect changes in unobservables and shopping needs. We address these threats to identification in several ways. First, we directly control for shopping needs X_{it} as in AH by including the log of the quantity index derived from the RPI as well as the log number of UPCs and product groups purchased per period. Second, we show that our results are robust to using a wide range of IVs, which assuages concerns that the IVs themselves may be reflecting changes in unobservables.

Second, we use proxies of labor demand shocks as IVs to further address threats to identification. We first document that based on the Current Population Survey Merged Outgoing Rotation Groups (CPS MORG), average hourly real wages have been rising over the sample period and unemployment rate rose substantially during the recession but gradually dropped to levels below those before the recession. This is consistent with the fall in shopping trips shown in Figure 4, which stagnated during the recession before falling further during the recovery. We show the trends for average wages and unemployment rate by gender and age group in Appendix Figure G18, and by gender and education group in Appendix Figure G19. As seen in these figures, trends do differ by gender, age, and education.

We follow Aguiar et al. (2013) and Aguiar et al. (2021) by using these differential changes as proxies of labor demand shocks that affect the opportunity cost of time, using national-level changes in both average wages and unemployment rates by household group as IVs. Our identifying assumption is that these changes reflect differential shifts in labor demand across various household demographic characteristics that are plausibly exogenous to unobservables affecting individual household shopping behavior. We construct these proxies at various levels of aggregation across household characteristics in the CPS MORG, such as age and education cells across the US. We match these proxies to households in the HMS using the mode of their characteristics over the sample period to avoid using variation induced by shifts in characteristics over time.

4.3.2 Results

We first show the cumulative effect of non-employment on household HHI and the number of shopping trips for both female and male heads in Figure 9 and Appendix Figure G20 respectively, using both OLS and following DCDH. We find that pre-trends are parallel and non-employment for female heads leads to a decline in household HHI by about 1 percentage point and a rise in the number of shopping trips by about 5%, while results are statistically insignificant for male heads. This is consistent with D'Acunto et al. (2021), who document that females are predominantly the primary shopper in households with both genders as household heads. These figures suggest that shifts in the opportunity cost of time of primary shoppers of households can have a significant effect on household HHI and shopping time. Given that the female employment rate for prime-aged workers rose from 69% to 74% during the recovery in our sample period as shown in Appendix Figure G17, we calculate a BOTE of close to 1% from the rise in prime-aged female employment rates.[18] Since these estimates are only relevant for the extensive margin, we further investigate the impact of changes in average real wages and unemployment rates using our proxies for labor demand shocks.

Next, we show our results from estimating equation (8) using additional household characteristics as IVs in Table 3. We first show OLS results with and without controls for shopping needs. We find that adding controls increases the effect of shopping trips on household HHI. A 1% decrease in the number of trips increases household HHI by about 0.2 percentage points. This is consistent with the fact that unobservables affecting household HHI, such as changes in shopping needs, may also affect the number of shopping trips. For the remaining specifications, we continue to control for shopping needs to isolate the effect of time costs on household HHI separate from the effect it also has on shopping quantities.

We find that using household age, size, and employment as IVs generate nearly identical coefficients in line with our event study estimates. These results are also robust to adding a county-time fixed effect to control for potential supply-side changes that vary over time within a county. The size of the first-stage F-stat implies the IV is relevant and the signs of the first-stage coefficients are consistent with a priori reasoning, i.e. increased household age, decreased household size, and unemployment all increase the amount of shopping trips. Using the average change in shopping trips over the sample period to calculate our BOTE, we find that shopping time has the potential to explain the entire rise in household HHI, with BOTEs ranging from 40%-100%. However, this BOTE can reflect both endogenous and exogenous factors affecting shopping time.

[18] About 50% of households have a prime-aged female head so a 5 percentage point increase in female employment rates can raise household HHI by 0.5*0.05*0.01 = 0.025 percentage points, which is about 0.8% of the 3 percentage point rise in household HHI during the recovery.

Therefore, we now focus on a potentially exogenous force decreasing shopping time by presenting results in Table 4 using national changes in group-level averages of male and female real wages, unemployment rates, and both as IVs. These groups include (1) age, (2) education, and (3) age plus education combined. While the magnitudes of the estimated coefficients do vary by IV, nearly all coefficients are statistically and economically significant. In all of the specifications, the first-stage F-stat passes the usual thresholds and the signs of the first-stage coefficient are consistent with a priori reasoning, i.e. rises in wages and drops in unemployment decrease shopping time. The BOTE using changes in shopping trips range from about 80%-220%, while the BOTE using only changes in the IVs range from about 0-14%. This suggests that labor demand shocks by age and education group do increase the opportunity cost of time and can explain up to 14% of the rise in household HHI.

We show that results are qualitatively similar when adding county-time fixed effects in Appendix Table F14. We also show results using other measures of concentration and varieties per trip in Appendix Table F15 and F16. We find that decreases in shopping time increases the number of retailers visited, increases household-category HHI, decreases cross-category variance, increases cross-household variance, and increases varieties per trip, all consistent with the trends we see in Section 3.

4.4 Other Hypotheses

What are some other hypotheses that can explain the rise in household concentration? We discuss this in detail in Appendix Section C. We suggest that changes in the cost of travel, improvements in leisure technology, diverging tastes between households, and increases in consumer loyalty could be contributing factors, whereas the rise of online shopping, converging tastes within households due to positive assortative matching in household formation, income inequality, and income polarization are less likely to be key contributing factors. We leave an in-depth exploration of these other factors for future research.

5 Model

In this section, we develop a model to rationalize our results and highlight the implications for market power and welfare. We model the demand-side as a two-layer nested CES utility function following Hottman et al. (2016) among others. The first nest contains firms and the second nest contains products, which enables consumers to choose from any number of multi-product firms.[19] We then introduce a heterogeneous cost to visiting each firm following

[19] We use the terms consumers and households interchangeably.

Bronnenberg (2015) and allow consumers to derive utility from leisure with a Cobb-Douglas consumption-leisure utility function. Given the model, we derive comparative statics for each of the variables of interest, showing the conditions under which evolutions in both supply and demand can increase household concentration and highlight the resulting welfare implications. We then calibrate the model to quantify the effect of changes in household concentration on market power. All derivations are shown in Appendix Section E.

5.1 Demand

Consumers have a Cobb-Douglas utility function and derive utility from (1) quantities of a composite consumption good $X(\mathcal{V})$, (2) the total variety of these goods $\mathcal{V}$, and (3) leisure $L(\mathcal{V})$, with a preference for leisure of ρ as shown in equation (9).[20] They maximize utility by choosing $X(\mathcal{V})$, $\mathcal{V}$, $L(\mathcal{V})$, and labor supply h, subject to a time constraint and budget constraint. The amount of time they have is T, which they split between shopping time $\tau(\mathcal{V})$, leisure L, and labor supply h. Shopping time $\tau(\mathcal{V})$ is the integral of the shopping cost for each variety $\mu(\nu)$ over all varieties $\mathcal{V}$. Consumers have income Y from supplying labor h at wage w and non-labor income K, which is used to buy consumption goods with price $p(\nu)$ and quantities $x(\nu)$ for each variety ν. We can combine the time and budget constraints into the full income constraint in equation (10).

$$
\begin{aligned}
&\max_{X,\mathcal{V},L,h} U(X(\mathcal{V}),L(\mathcal{V})) = X(\mathcal{V})^{1-\rho}L(\mathcal{V})^{\rho} && (9)\\
&s.t. \quad \tau(\mathcal{V}) + L + h = T, \text{ where } \tau(\mathcal{V}) = \int_{\nu\in\mathcal{V}} \mu(\nu)d\nu &&\\
&s.t.\ Y = wh + K = \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu = PX(\mathcal{V}) &&\\
&s.t.\ wT + K = \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu + w(\tau(\mathcal{V}) + L) && (10)
\end{aligned}
$$

Consumption good $X(\mathcal{V})$ contains two CES nests for firms and products as shown in equation (11). The first nest consists of firms ν, which are retailers in our context. Each firm has a taste parameter $\varphi^F(\nu)$ and quantity consumed $x^F(\nu)$ with an elasticity of substitution σ^F between firms. The second nest contains products u, which enables consumers to choose a set of products $\mathcal{U}_f$ from firm f, with consumers choosing the set of multiproduct firms $\mathcal{V}$.[21] Each product has a taste parameter $\varphi^U(u)$ and quantity consumed $x^U(u)$ with an elasticity

[20]Results are nearly identical and give the same intuition using a separable consumption-leisure constant relative risk aversion (CRRA) utility function that is common in labor supply life-cycle models.

[21]To maintain consistency with previous literature, we use ν and f interchangeably for firms.

of substitution σ^U between products.

$$X(\mathcal{V}) = \left(\int_{\nu\in\mathcal{V}} \left(\varphi^F(\nu)x^F(\nu)\right)^{\frac{\sigma^F-1}{\sigma^F}} d\nu\right)^{\frac{\sigma^F}{\sigma^F-1}},\ x^F(\mathcal{U}_f) = \left(\int_{u\in\mathcal{U}_f} \left(\varphi^U(u)x^U(u)\right)^{\frac{\sigma^U-1}{\sigma^U}} du\right)^{\frac{\sigma^U}{\sigma^U-1}} \tag{11}$$

Solving for the demand at the firm and product level following Hottman et al. (2016), denoted as $x^F(\nu)$ and $x^U(u)$ respectively, we have the following equations:

$$x^F(\nu) = A(\mathcal{V})p^F(\nu)^{-\sigma^F}\varphi^F(\nu)^{\sigma^F-1} \tag{12}$$

$$A(\mathcal{V}) = YP^{\sigma^F-1},\ P = \left(\int_{\nu\in\mathcal{V}} \left(\frac{p^F(\nu)}{\varphi^F(\nu)}\right)^{1-\sigma^F} d\nu\right)^{\frac{1}{1-\sigma^F}}$$

$$x^U(u) = x^F(\nu)(\varphi^U_u)^{\sigma^U-1}\left(\frac{P^F_f}{P^U_u}\right)^{\sigma^U} \tag{13}$$

$$\begin{aligned} P^F_f = p^F(\nu) &= \left(\int_{u\in\mathcal{U}_f} \left(\frac{P^U_u}{\varphi^U_u}\right)^{1-\sigma^U} du\right)^{\frac{1}{1-\sigma^U}} \\ &= \left(\underbrace{N_f}_{\text{Scope}} \underbrace{\frac{1}{N_f}\int_{u\in\mathcal{U}_f} \left(\frac{P^U_u}{\varphi^U_u}\right)^{1-\sigma^U} du}_{\text{Average product taste-adjusted prices}}\right)^{\frac{1}{1-\sigma^U}} \end{aligned} \tag{14}$$

The price index for each firm, which we denote interchangeably as P^F_f and $p^F(\nu)$, can be written as its scope N_f multiplied by the average product taste-adjusted prices, where P^U_u is the price for product u and φ^U_u is the idiosyncratic taste for product u.[22] Next, we can use the first-order condition (FOC) to derive the optimal composite good consumed and leisure as follows:

$$L(\mathcal{V}) = \frac{\rho\left(w\left(T-\tau(\mathcal{V})\right)+K\right)}{w} \tag{15}$$

$$X(\mathcal{V}) = \frac{(1-\rho)\left(w\left(T-\tau(\mathcal{V})\right)+K\right)}{P} = \frac{Y}{P} = \frac{w(T-\tau(\mathcal{V})-L)+K}{P} \tag{16}$$

[22]We could further add a cost per product to generate zero purchases for certain products. However, we abstract away from this feature for tractability.

Substituting these expressions into the utility function, we can follow Bronnenberg (2015) and derive that the optimal cutoff variety ν_D in the set of varieties D satisfies the following condition:

$$\frac{A(D)\left(\frac{p(\nu_D)}{\varphi(\nu_D)}\right)^{1-\sigma^F}}{\sigma^F-1} - w\mu(\nu_D) = 0 \tag{17}$$
$$A(D) = YP^{\sigma^F-1} = (1-\rho)\left(w\left(T-\tau(D)\right)+K\right)\left(\int_{\nu\in D}\left(\frac{p^F(\nu)}{\varphi^F(\nu)}\right)^{1-\sigma^F} d\nu\right)^{-1}$$

Equation 17 equalizes the marginal benefit and cost of each additional variety. The marginal benefit increases as the taste-adjusted price of firm f decreases relative to the price index P the household faces, decreases with elasticity of substitution σ^F, and increases with income Y. The marginal cost increases with wage w, which also represents the opportunity cost of time, and shopping cost μ. Each additional variety will be consumed whenever the marginal benefit exceeds the marginal cost.

Next, we assume that firms have identical taste parameters φ^F equal to one. This is simply done for expositional purposes and allows us to index each firm by their price index rather their taste-adjusted price index $\frac{P_f^F}{\varphi_f^F}$. We also assume that P_f^F lies on a continuum $[\underline{P^F}, \overline{P^F}]$.[23] We assume that the shopping cost at each firm is a function of the price index P_f^F in each firm, with $\mu'(P_f^F) < 0$. This is consistent with the fact that firms with larger variety and lower prices, i.e. lower P_f^F, tend to be located farther away from consumers relative to firms with lower variety and higher prices, due to higher costs of land in areas with higher population density and the need for stores with larger square footage in order to stock higher variety and lower prices through economies of scale. We can show using equation (17) that if $\mu'(P_f^F)$ is small enough in absolute value, the net marginal gain of shopping at an additional firm is monotonically decreasing in P_f^F, and there exists a unique cutoff firm P^{F*} that satisfies the following condition, such that consumers only buy from firms that have a price index within the set $[\underline{P^F}, P^{F*}]$:

$$\left(P^{F*}\right)^{1-\sigma^F}\frac{1}{\int_{\underline{P^F}}^{P^{F*}}\left(P_f^F\right)^{1-\sigma^F} dP_f^F}\frac{(1-\rho)\left(w\left(T-\int_{\underline{P^F}}^{P^{F*}}\mu(p)dp\right)+K\right)}{w(\sigma^F-1)} - \mu(P^{F*}) = 0 \tag{18}$$

[23]Assuming the firms lie on a continuum indexed by P_f^F allows for higher analytical tractability and the use of integrals. We can also allow for discrete P_f^F and use summations instead with similar intuitions for all of our derivations.

5.2 Household Retail Concentration

Given our derivations, the market share for household i for each firm f can be written as follows:

$$S^F_{fi} = \frac{P^F_f x^F(\nu)}{Y} = \begin{cases} \left(P \frac{\varphi^F_f}{P^F_f}\right)^{\sigma^F - 1} & \text{if } \frac{Y P^{\sigma^F-1}\left(\frac{P^F_f}{\varphi^F_f}\right)^{1-\sigma^F}}{w(\sigma^F-1)} \geq \mu(P^F_f) \\ 0 & \text{otherwise} \end{cases}$$

Given our simplifying assumptions, we can further simplify the market share expression and write the household retail HHI for household i as follows:

$$H_i = \int_f \left(S^F_{fi}\right)^2 df = \frac{\int_{\underline{P^F}}^{P^{F*}} \left(P^F_f\right)^{2\left(1-\sigma^F\right)} dP^F_f}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P^F_f\right)^{1-\sigma^F} dP^F_f\right)^2} \tag{19}$$

We can then show some comparative statics of household retail HHI in response to changes in various parameters. First, we show that any parameters, in this case denoted by t, that decrease the cutoff firm price index P^{F*} will increase household retail HHI. This is intuitive since the set of retailers that households consume from $[\underline{P^F}, P^{F*}]$ will decrease in size as P^{F*} decreases.

$$\frac{dP^{F*}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0 \tag{20}$$

Given this fact, we can use equation (18) to derive comparative statics of the cutoff firm price index P^{F*} in response to changes in various parameters to see how these parameters change household retail HHI. These comparative statics can then be compared against our empirical findings. We summarize our the implications of our theory for household concentration and welfare in Table 5.

First, consider changes on the demand side. Let the shopping cost be $\mu(P^F) = t\delta(P^F)$, where t is the time cost per distance traveled and $\delta(P^F)$ is the distance of the consumer from each firm indexed by P^F. We derive that if the condition required for the existence of a unique cutoff price index holds, as wage w increases, household retail HHI also increases. As the time cost per distance traveled t increases, household retail HHI increases. Let the preference for leisure be written as $\theta\rho$, where θ is a form of leisure technology. By varying θ, we can keep the preference for consumption unchanged. This would be consistent with increases in leisure technology highlighted in Aguiar et al. (2021). We find that as leisure technology θ increases, household retail HHI increases.

All of these results are intuitive and consistent with our empirical results that a rise in the opportunity cost of time raises household retail HHI. As the shopping cost increases or the relative cost of leisure decreases, households spend less time shopping and visit fewer retailers, raising household retail HHI.

Second, consider changes on the supply side such as the opening of new firms, which decrease μ. Denoting such changes as an increase in β, we can write $\mu = \mu(P^F, \beta)$ and $\frac{\partial \mu}{\partial \beta} < 0$. A rise in β generates the opposite effect as a rise in t, decreasing household retail HHI. This is because households can now spend more time shopping and visit more retailers, decreasing household retail HHI. This is consistent with our empirical results that the opening of club stores, which provide similar amounts of product variety to many other stores, decreases household retail HHI. Alternatively, consider the introduction of supercenters, which we model as a fall in $\underline{P^F}$, the lower limit for the range of firm price indices, since supercenters represent a shopping format that provides an unprecedented number of products at low prices in a single store. We find that a decrease in $\underline{P^F}$ raises household retail HHI if there is a sufficient decrease in P^{F*} as shown in Appendix Section E, consistent with our empirical results. Intuitively, a fall in $\underline{P^F}$ represents the introduction of a new firm that lowers the overall price index, increasing the relative price of the initial cutoff price index P^{F*} such that P^{F*} falls.

Third, consider changes on the supply side such as increasing economies of scale or scope, which decrease P^F by lowering prices or raising variety. Denoting such changes as an increase in α, we can write $P^F = P^F(\alpha)$ and $\frac{dP^F}{d\alpha} < 0$. A rise in α could either increase or decrease household retail HHI, depending on whether the firm price index decreases favor firms with a larger or smaller initial firm price index. Let $f(P^F, \alpha) = \left(P^F(\alpha)\right)^{1-\sigma^F}$. We derive a condition under which household retail HHI would increase as α increases, which can be interpreted as economies of scale or scope that disproportionately favor firms with lower prices and larger variety, consistent with our empirical results in Appendix Table F9.

$$\int f \frac{df}{d\alpha} > \int f \int \frac{df}{d\alpha} \Rightarrow \frac{dH_i}{d\alpha} > 0 \tag{21}$$

5.3 Welfare

Substituting the optimal quantity consumed X, variety $\mathcal{V}$, and leisure L into the utility, we can derive consumer welfare in this model. Differentiating utility with respect to various parameters in our model and using the envelope theorem, we can show how welfare changes in response to various changes on the demand side and supply side.

First, consider changes on the demand side. A rise in the time cost per distance traveled

t lowers welfare, since more time is spent shopping, decreasing both income Y and leisure L. A rise in wage w raises welfare as it increases income Y. A rise in leisure technology θ unambiguously raises welfare by increasing the value of leisure. Therefore, the source of the rise in opportunity cost of time is crucial in determining whether welfare increases or decreases.

Second, consider changes on the supply side. A rise in β which decreases shopping cost μ increases welfare in the same way as a reduction in t. A fall in minimum firm price index $\underline{P^F}$ increases welfare. A rise in α which decreases price indices within firms increases welfare. Therefore, all of these supply-side changes should increase welfare.

5.4 Heterogeneity

To account for changes in cross-household variance, we need to allow for heterogeneity across consumers. To model such heterogeneity in the simplest way possible, we first allow for two groups of consumers. The first group of consumers only buys from firms that provide firm price indices below the cutoff firm price index $\overline{P^{F*}}$ that we previously used, consuming from the set of firms $[\underline{P^F}, \overline{P^{F*}}]$. We introduce a second group of consumers that only buys from firms that offer a price index above the cutoff firm price index $\underline{P^{F*}}$, consuming from the set of firms $[\underline{P^{F*}}, \overline{P^F}]$. We find that as long as $\mu'(P^F)$ is large enough in absolute value, the net marginal gain of shopping at an additional firm is monotonically increasing in P^F_f, and there exists such a unique cutoff firm $\underline{P^{F*}}$. We can consider such consumers as those who find it far more costly to visit firms with large variety and low prices that are generally located away from population centers, e.g. consumers that do not have cars.

We then derive comparative statistics on household retail concentration for this second group of consumers. All of the derivations have the same intuition. First, household retail HHI increases as the number of retailers visited decreases, which in this case corresponds to an increase in $\underline{P^{F*}}$. Second, the results on the comparative statistics remain unchanged. A decrease in $\underline{P^{F*}}$ now corresponds to an increase in $\underline{P^{F*}}$, both of which increase household retail HHI.

How does the introduction of this second group of consumers account for a rise in cross-household variance? If $\underline{P^{F*}} < \overline{P^{F*}}$, this implies that there are a set of firms that both groups will consume from. Demand and supply-side changes that increase household retail HHI will decrease $\overline{P^{F*}}$ for one group and increase $\underline{P^{F*}}$ for the other group, decreasing the set of firms $[\underline{P^{F*}}, \overline{P^{F*}}]$ that both groups consume from. In other words, the first group will polarize towards low price index firms with larger variety, lower prices, and higher shopping costs, while the second group will polarize towards high price index firms with low variety, higher prices, and lower shopping costs, increasing cross-household variance. Allowing for a continuum of consumer groups gives the same intuition, in which households may polarize towards their preferred firms when shrinking the set of firms they consume from.[24]

Therefore, we find that demand-side changes such as a rise in w, θ, and t will all increase household retail HHI and increase cross-household variance by decreasing $\overline{P^{F*}}$ and increasing $\underline{P^{F*}}$. For supply-side changes, a rise in β will affect $\overline{P^{F*}}$ and $\underline{P^{F*}}$ as long as it affects the set of firms consumers visit. For example, if the second group of consumers does not visit club stores, then the entry of club stores do not change their household retail HHI, and any change in cross-household variance will only result from changes in consumption of the first group, i.e. an increase in $\overline{P^{F*}}$ and not a drop in $\underline{P^{F*}}$. Likewise, a fall in $\underline{P^{F}}$ will not affect $\underline{P^{F*}}$ as long as $\underline{P^{F*}} > \underline{P^{F}}$, so any change in cross-household variance will result from a fall in $\overline{P^{F*}}$ and not a rise in $\underline{P^{F*}}$. Hence, these changes may have a smaller effect on cross-household variance, consistent with our findings that the entry of supercenters and club stores do not have a statistically significant effect on cross-household variance. Likewise, whether an increase in α increases cross-household variance also depends on which group of firms are disproportionately changing their price indices.[25]

As for changes in welfare, our previous derivations are general and allow for consumers that belong to both the first and second group. One exception is that a fall in $\underline{P^{F}}$ should have no effect on households that do not consume from these firms.

5.5 Markups

What are the implications of our empirical results for market power? First, consider a general profit maximization problem for each firm f:[26]

$$\max_{P_f^F} \Pi_f^F = S_f^F \int_i Y_i di - C_f$$

Firm f chooses its firm price index P_f^F given its cost function C_f and its revenue is driven

[24]We can further allow households to consume from a set of firms $[\underline{P^{F*}}, \overline{P^{F*}}]$, i.e. their preferred set may not contain the lower or upper limits, by assuming that the second derivative of the cutoff equation is negative, which holds when the shopping cost function is sufficiently convex. In this case, we would not need to assume the net marginal gain from shopping at an additional firm is monotonically increasing or decreasing in P_f^F.

[25]We can further explain the rise in one-stop shopping if consumers predominantly shift towards high-variety firms such as supercenters with a fall in $\underline{P^{F*}}$. One way to allow them to increase varieties per trip when polarizing towards low price index firms when increasing $\overline{P^{F*}}$ would require introducing costs per variety, which we abstract away from for tractability as mentioned.

[26]We show a more detailed supply-side setup that generates these equations in Appendix Section E.5.

by demand equations that determine its market share S_f^F. The FOC is then given by:

$$\frac{\partial \Pi_f^F}{\partial P_f^F} = \int_i Y_i di \frac{\partial S_f^F}{\partial P_f^F} - \frac{\partial C_f}{\partial P_f^F} = 0$$

Next, we have the following identity which relates the elasticity of the firm's residual demand elasticity of demand ε_f^F to the market share elasticity with respect to price:

$$\varepsilon_f^F = 1 - \frac{\partial S_f^F}{\partial P_f^F} \frac{P_f^F}{S_f^F}$$

The FOC then gives us the following equation for the firm-level markup $\mathcal{M}_f^F$ for firm f, which is defined as the price index P_f^F divided by the marginal cost MC_f, as a function of ε_f^F, which is the same across all products.[27]

$$\mathcal{M}_f^F \equiv \frac{P_f^F}{MC_f} = \frac{\varepsilon_f^F}{\varepsilon_f^F - 1}$$

We then show how the firm market share for the entire market S_f^F is related to the individual household market share s_{fi}^F. Recall that the market share S_{fi}^F of firm f for household i can be written as follows:

$$S_{fi}^F = S_f^F(D_i) = \begin{cases} \left(P \frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F - 1} & \text{if } \frac{Y S_{fi}^F}{w(\sigma^F - 1)} \geq \mu_{fi}^F = \mu_i(P_f^F) \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

$$P = P(D_i) = \left(\int_{f \in D_i} \left(\frac{P_f^F}{\varphi_f^F} \right)^{1-\sigma^F} dP_f^F \right)^{\frac{1}{1-\sigma^F}}$$

$$Y_i = Y(D_i) = (1-\rho)\left(w \left(T - \int_{f \in D_i} \mu_{fi}^F df \right) + K \right)$$

The market share, household price index, and income are all dependent on the specific set of firms D_i that household i consumes from, and D_i is a function of μ_{fi}^F by equation (17). We now allow for consumer heterogeneity more generally, allowing each household i to have a different shopping cost function μ_{fi}^F with probability density function $f(\mu_{fi}^F)$, where the shopping cost function is decreasing in P_f^F as mentioned previously. Since equation (17) is monotonic in μ_{fi}^F, for each firm f there exists a cutoff household $i*$ such that equation (17)

[27] This is true even without imposing product symmetry and is a property of nested demand systems, as mentioned in Hottman et al. (2016).

holds. Household $i*$ then represents the marginal consumer for firm f. We can then write the market share S_f^F as follows:

$$\forall f,\ \exists i^* \text{ s.t. } \frac{Y(D_{i*})S_{fi}^F(D_{i*})}{\sigma^F - 1} = w\mu_{fi*}^F \tag{23}$$
$$S_f^F = \frac{\int_i Y_i S_{fi}^F di}{\int_i Y_i di} = \frac{\int_0^{\mu_{fi*}^F} Y(D_i)S_f^F(D_i)f(\mu_{fi}^F)d\mu_{fi}^F}{\int_i Y_i di}$$

For ease of exposition, assume that income Y_i is independent of P_f^F.[28] We can then rewrite the market share elasticity more simply as follows:

$$\begin{aligned}\frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F} &= \frac{\partial S_f^F(D_{i*})}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_{i*})}\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F \\ &+ \int_0^{\mu_{fi*}^F}\frac{\partial S_f^F(D_i)}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_i)}\frac{S_f^F(D_i)}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}d\mu_{fi}^F\end{aligned}$$

Given the household market share S_{fi}^F, we can derive the following household market share elasticities under monopolistic competition and Bertrand competition:

$$\frac{\partial S_{fi}^F}{\partial P_f^F}\frac{P_f^F}{S_{fi}^F} = \begin{cases} 1-\sigma^F & \text{if monopolistic competition} \\ (1-S_{fi}^F)(1-\sigma^F) & \text{if Bertrand competition}\end{cases}$$

Under monopolistic competition, the market share elasticity is only dependent on the elasticity of substitution σ^F. Under Bertrand competition, a change in the price index of firm f affects the aggregate price index P. The market share elasticity depends on both σ^F and the market share S_{fi}^F. A larger market share lowers the absolute value of the market share elasticity, lowering the elasticity of demand and raising the markup. Substituting the household market share elasticities into the market share elasticity and hence the demand elasticity, we can derive the markup under monopolistic competition and

[28] As shown above, relaxing this assumption would allow the Y_i to shift as weights in response to changes in prices, which complicates the analysis without affecting the main intuition.

Bertrand competition:

Monopolistic Competition

$$\mathcal{M}_f^F = \frac{\sigma^F + EM}{\sigma^F - 1 + EM} \tag{24}$$

$$EM = (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F = \frac{\left(S_f^F(D_{i*})\right)^2}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\frac{Y(D_{i*})}{w}$$

Bertrand Competition

$$\mathcal{M}_f^F = \frac{\sigma^F + EM - IM}{\sigma^F - 1 + EM - IM} \tag{25}$$

$$EM = (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F(1 - S_f^F(D_{i*})) \tag{26}$$

$$IM = (\sigma^F - 1)\int_0^{\mu_{fi*}^F} \frac{1}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}\left(S_f^F(D_i)\right)^2 d\mu_{fi}^F \tag{27}$$

Therefore, the markup differs from the standard CES markup $\frac{\sigma^F}{\sigma^F-1}$ in two ways. First, there is an additional extensive margin term EM that lowers the markup, since the existence of such marginal consumers increases the demand elasticity. This term is a function of several terms which include the share of firm sales from marginal consumers and the shopping cost at the margin. Note that there is also an additional term $1 - S_f^F(D_{i*})$ to account for the effect on the aggregate price index under Bertrand competition. This term is similar to the additional extensive margin term in Neiman and Vavra (2021). In contrast to their paper which focuses on taste heterogeneity across products, we allow for consumer heterogeneity to work through the shopping cost μ.

Second, there is an intensive margin term IM that increases markups due to Bertrand competition in addition to consumer heterogeneity. This term has also been highlighted in Feenstra et al. (2022), who again focus on taste heterogeneity across products. Intuitively, firms do not weight demand elasticities across consumers equally, but optimally use a greater weight on low-elasticity consumers, such that when consumer heterogeneity is present, they can charge higher markups and obtain higher profits, since the gains from charging higher markups to the lower elasticity consumers offsets the loss in demand on higher elasticity consumers. Crucially, this term increases as the variance of market shares across households increases, since firms can charge higher markups when there is a larger share of lower elasticity consumers.

The aggregate markup $\mathcal{M}$ is then the share-weighted average of all firm markups:

$$\mathcal{M} = \int_f S_f^F \mathcal{M}_f^F df \tag{28}$$

Therefore, how aggregate market power changes as a result of our empirical findings is ambiguous and requires a more detailed quantitative analysis, depending on the shopping cost distribution and market structure assumptions. As household retail concentration increases along with cross-household variance, there should be some firms with increased household market shares, generating potentially offsetting effects on markups by raising the EM term and lowering the consumer heterogeneity term IM, such that the effect on firms' markups, as well as the aggregate markup, is ambiguous. Therefore, we turn to a calibration to quantify these offsetting effects.

5.6 Calibration

To quantify the effect of changes in household concentration on markups, we now calibrate firm markups under Bertrand competition using equation (25) and aggregate to the national level using equation (28). We first focus on the IM term in equation (27). This term is a weighted-average of the household market share of each consumer in firm f, where the weights are the shares of expenditure accounted for by each household, multiplied by $\sigma^F - 1$. This term closely resembles the household HHI but aggregates to the firm level rather than the household level. We calibrate this formula using household-firm market shares by year in our data, and an estimate of σ^F equal to 4.5 following Hottman (2019).

Next, we also calibrate an upper bound for the EM term in equation (26). We first calculate an upper bound for shopping costs μ_{fi}^F by using equation (22). We aggregate this equation to each household-firm-year by summing over the number of shopping trips. In other words, the inferred upper bound for μ_{fi}^F is higher for a household i with high expenditures YS_{fi}^F in firm f but a low number of trips.[29] Next, we can obtain an upper bound on the density of consumers at the margin $f(\mu_{fi*}^F)$ by using the maximum density from the distribution of upper bounds for each firm.[30] Finally, we use the mean household expenditures and shares for each firm to approximate the remaining terms.[31]

[29] We use the average hourly wage w from the CPS by year, assume a constant 25% tax rate to obtain a post-tax hourly wage, then scale it by 0.75 following Goldszmidt et al. (2020) to obtain an estimate of the opportunity cost of time.

[30] We show in Appendix Figure G25 that the distribution closely follows a log-normal distribution, which we use to approximate the distribution for each firm. The magnitudes of the shopping cost upper bounds are also the same order of magnitude to Panel D in Figure 5 of Coibion et al. (2021), which shows that the average duration of a shopping trip is around 25 minutes, i.e. about 0.4 hours.

[31] We multiply the mean household expenditure and divide it by the total firm expenditure (both include

We show the calibrated aggregate markups in Figure 10. Aggregate markups rose from 1.46 to about 1.51 over the sample period. This trend closely resembles the trend in household HHI. We also calculate the markups using market shares aggregated to the national, market, and county level respectively.[32] We find that both the levels and changes in markups are much smaller when calculated at these higher levels of geographic aggregation, with the markups calculated using data aggregated to the county level giving the smallest increase. This is consistent with Figure 1, which shows that the increase in concentration is smallest at the county level.

Our results at the national and local level closely mirror those in Smith and Ocampo (2021), whose model implies that increases in local concentration can increase retailers' margins by at most up to 1.7 percentage points out of a 6 percentage point increase in retailer gross margins based on the Annual Retail Trade Survey from 1992-2012. Our results imply that inferring the level of competition from changes to outcomes at the geographic-market level, even if defined at a low level of aggregation, could give different results from looking at household outcomes and mask changes in household heterogeneity. The intuition is that under Bertrand competition, a rise in household concentration implies a lower demand elasticity for each household. Given that firms price based on a weighted-average of household demand elasticities, this allows firms to charge higher markups.

We further investigate how changes in weights are driving the aggregate markup in Figure 11. We plot the weighted average of firm markups by fixing the weights to the total market share of each firm across the sample period using only firms that did not experience entry or exit, as well as the simple average of firm markups, again for firms that were present in the entire sample period. This allows us to separate the contribution of shifts in market shares towards high-markup firms from the rise in markups within firm. We find that the change in weighted markups is about 50% of the change in aggregate markup, which implies that both within-firm changes and reallocation of market shares towards high-markup firms are responsible for the rise in aggregate markups. The simple average of markups is actually falling, which is consistent with high-markup firms taking up a larger proportion of market shares.[33]

We show in Appendix Figure G27 that our results imply that the extensive margin contributes a negligible proportion to both levels and changes in markups. This is because

projection factors) to obtain $Y(D_{i*})S_f^F(D_{i*})$ and $S_f^F \int_i Y_i di$ respectively. We use the mean household market share for firm f to approximate $S_f^F(D_{i*})$.

[32]For these measures we only calculate the IM term, since the EM term can only be defined at the household level.

[33]Note that the average markup is similar to those in Hottman (2019), given that we use the same elasticity of substitution.

the density of shopping costs is continuous so the proportion of consumers at the margin is minuscule. Even if we allowed the proportion of expenditures accounted for by extensive margin households to increase to up to 25%, the aggregate markup still increases by close to 4 percentage points over the sample period. This implies that the variation in markups under monopolistic competition would be fairly small in our context.

We also show the markups for the top 10 firms by revenue in Appendix Figure G28. There is a considerable amount of dispersion, as firms achieve high revenue nationally not only by having higher household market shares, which leads to higher markups, but also by having many consumers. Nonetheless, we can see that markups are increasing for 8 out of the 10 top firms.[34]

How do these changes in markups affect different households? We calculate markups for each household by multiplying their expenditure share in each firm with that firm's markup. We do this both by allowing expenditure shares to vary over time or fixing them with shares calculated across the entire sample period. We show in Appendix Figure G29 and Appendix Figure G30 the household markups by percentile for both varying and fixed weights, in levels and changes since 2004. We find that similar to Appendix Figure G2c, households with higher levels of markups are experiencing larger increases. This implies that inequality in markups across households has increased over time.

Overall, our results imply that our model implies that the forces that raise household concentration, such as the rise of supercenters, product variety, and opportunity costs of time, should increase welfare, but some of these welfare gains will be offset by a rise in markups.[35]

[34] We also show that our results are robust to assuming Cournot rather than Bertrand competition in Appendix Figure G26, although the aggregate markup is much higher than those from the previous literature, which suggests that Bertrand competition may be a more reasonable assumption.

[35] While we can quantify the effect on aggregate markups, we refer to previous literature that attempts to quantify the welfare gains of the various supply- and demand-side factors we analyze. For example, on the supply side, Atkin et al. (2017) and Leung and Li (2022) both find that big-box stores, in particular supercenters, offer more variety and lower prices, which increases consumer welfare. As a result, the entry of more big-box retailers, combined with growth in variety within existing stores, are likely to be beneficial to consumers. On the demand side, Thomassen et al. (2017) use a multi-category multi-seller demand model estimated using UK consumer data to show that consumers inclined to one-stop shopping, as opposed to multi-stop shopping, have a greater pro-competitive impact because although they have lower demand elasticities, they also generate relatively large complementary cross-category effects. Since stores internalize these cross-price elasticities, market power can be reduced substantially as a result. Hence, whether market power is reduced and consumer welfare is increased hinges on the relative sizes of the changes in the cross-category effect against the fall in demand elasticities from increased one-stop shopping. We abstract away from these cross-category effects but we also highlight the importance of both the intensive and extensive margin in determining markups.

6 Conclusion

In this paper, we document that household concentration has risen substantially in the US retail sector from 2004-2019. Households do more one-stop shopping and increasingly shop at different retailers from one another, even as the number of local retail establishments has increased. This is consistent with a negligible change in regional concentration and a rise in national concentration, as regions are becoming more similar in their market share distributions.

We explain these facts using evolutions in supply and demand in the retail sector. On the supply side, we find that the number of big-box stores, namely supercenters and club stores, have increased. Utilizing an event study approach to estimate the impact of entry of these superstar retailers, we find that supercenters raise household HHI and one-stop shopping, while club stores decrease household HHI and one-stop shopping. This can be attributed to the fact that households buy far more varieties and pay lower prices in supercenters than any other channel type, including club stores. We then utilize an IV strategy to estimate the effect of changes in product variety and prices within existing stores. We find that increases in variety and lower prices increases household HHI. These supply-side changes by retailers can explain up to about 20% of the rise in household HHI.

On the demand side, we find that a fall in the number of shopping trips has the potential to explain the entire rise in household HHI using multiple IV strategies. One key factor that decreases the number of shopping trips is the rise in the opportunity cost of time, which leads to higher household HHI and increased one-stop shopping. We find that national changes in wages and unemployment rates can explain up to 14% of the rise in household HHI.

We develop a model to rationalize these results, and highlight the resulting implications for market power and welfare. While these supply-side changes that increase household concentration tend to be associated with increased welfare, the demand-side changes that increase household concentration have more ambiguous effects, and depend crucially on the source of these changes. Wage increases are beneficial for welfare, but increases in the cost of travel will lower welfare. Crucially, we show that household concentration, rather than market concentration, is tightly linked to markups. Our calibrated model suggests that the rise in household concentration led to a 5 percentage point increase in aggregate markups. Therefore, these supply- and demand-side developments that increase household concentration and welfare can also increase markups, offsetting some of these welfare gains. Given a rise in the availability of matched consumer-firm data from sources such as scanner data, bank account data, and credit data, we believe that household concentration can serve as an additional metric beyond market concentration to measure competition.

References

Aguiar, M., M. Bils, K. K. Charles, and E. Hurst (2021, February). Leisure Luxuries and the Labor Supply of Young Men. *Journal of Political Economy 129*(2), 337–382. Publisher: The University of Chicago Press.

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

Aguiar, M., E. Hurst, and L. Karabarbounis (2013, August). Time Use During the Great Recession. *American Economic Review 103*(5), 1664–1696.

Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019, November). Food Deserts and the Causes of Nutritional Inequality. *The Quarterly Journal of Economics 134*(4), 1793–1844.

Arcidiacono, P., P. B. Ellickson, C. F. Mela, and J. D. Singleton (2020, July). The Competitive Effects of Entry: Evidence from Supercenter Expansion. *American Economic Journal: Applied Economics 12*(3), 175–206.

Atkin, D., B. Faber, and M. Gonzalez-Navarro (2017, October). Retail Globalization and Household Welfare: Evidence from Mexico. *Journal of Political Economy 126*(1), 1–73.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020, May). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Balto, D. A. (2001, April). Supermarket Merger Enforcement. *Journal of Public Policy & Marketing 20*(1), 38–50. Publisher: SAGE Publications Inc.

Basker, E., S. Klimek, and P. H. Van (2012). Supersize It: The Growth of Retail Chains and the Rise of the "Big-Box" Store. *Journal of Economics & Management Strategy 21*(3), 541–582. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1530-9134.2012.00339.x.

Benkard, C. L., A. Yurukoglu, and A. L. Zhang (2021, April). Concentration in Product Markets. Working Paper 28745, National Bureau of Economic Research. Series: Working Paper Series.

Berry, S., M. Gaynor, and F. Scott Morton (2019, August). Do Increasing Markups Matter? Lessons from Empirical Industrial Organization. *Journal of Economic Perspectives 33*(3), 44–68.

Borgen, N. T., A. Haupt, and Ø. N. Wiborg (2021, April). A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model.

Bronnenberg, B. J. (2015). The provision of convenience and variety by the market. *The RAND Journal of Economics 46*(3), 480–498. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12094.

Bronnenberg, B. J. (2018, August). Retailing and consumer demand for convenience. *Handbook of Research on Retailing*. ISBN: 9781786430281 Publisher: Edward Elgar Publishing Section: Handbook of Research on Retailing.

Bronnenberg, B. J., T. J. Klein, and Y. Xu (2020). Consumer Time Budgets and Grocery Shopping Behavior. Working Paper.

Carden, A. and C. Courtemanche (2016, January). The evolution and impact of the general merchandise sector. *Handbook on the Economics of Retailing and Distribution*. ISBN: 9781783477388 Publisher: Edward Elgar Publishing Section: Handbook on the Economics of Retailing and Distribution.

Chernozhukov, V., I. Fernández-Val, and Y. Luo (2018). The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages. *Econometrica 86*(6), 1911–1938.

Coibion, O., Y. Gorodnichenko, and D. Koustas (2021, October). Consumption Inequality and the Frequency of Purchases. *American Economic Journal: Macroeconomics 13*(4), 449–482.

Covarrubias, M., G. Gutiérrez, and T. Philippon (2020, January). From Good to Bad Concentration? US Industries over the Past 30 Years. *NBER Macroeconomics Annual 34*, 1–46. Publisher: The University of Chicago Press.

de Chaisemartin, C. and X. D'Haultfœuille (2020, September). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–2996.

De Loecker, J., J. Eeckhout, and G. Unger (2020, May). 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.

Döpper, H., A. MacKay, N. Miller, and J. Stiebale (2021). Rising Markups and the Role of Consumer Preferences. Working Paper.

D'Acunto, F., U. Malmendier, and M. Weber (2021, May). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences 118*(21), e2008534118.

Ellickson, P. B. (2016, January). The evolution of the supermarket industry: from A & P to Walmart. *Handbook on the Economics of Retailing and Distribution*. ISBN: 9781783477388 Publisher: Edward Elgar Publishing Section: Handbook on the Economics of Retailing and Distribution.

Feenstra, R. C., L. Macedoni, and M. Xu (2022, January). Large Firms, Consumer Heterogeneity and the Rising Share of Profits. Working Paper 29646, National Bureau of Economic Research. Series: Working Paper Series.

Ganapati, S. (2021, August). Growing Oligopolies, Prices, Output, and Productivity. *American Economic Journal: Microeconomics 13*(3), 309–327.

Gelman, M., Y. Gorodnichenko, S. Kariv, D. Koustas, M. D. Shapiro, D. Silverman, and S. Tadelis (2016, December). The Response of Consumer Spending to Changes in Gasoline Prices. Working Paper 22969, National Bureau of Economic Research.

Goldszmidt, A., J. A. List, R. D. Metcalfe, I. Muir, V. K. Smith, and J. Wang (2020, December). The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments. Technical Report w28208, National Bureau of Economic Research.

Grullon, G., Y. Larkin, and R. Michaely (2019, July). Are US Industries Becoming More Concentrated? *Review of Finance 23*(4), 697–743.

Hausman, J. A. and T. F. Bresnahan (2008, April). 5. Valuation of New Goods under Perfect and Imperfect Competition. In *The Economics of New Goods*, pp. 209–248. University of Chicago Press.

Hortaçsu, A. and C. Syverson (2015). The ongoing evolution of us retail: A format tug-of-war. *Journal of Economic Perspectives 29*(4), 89–112.

Hottman, C. (2019). Retail Markups, Misallocation, and Store Variety in the US. Working Paper.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016, August). Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics 131*(3), 1291–1364.

Hsieh, C.-T. and E. Rossi-Hansberg (2019, June). The Industrial Revolution in Services. Working Paper 25968, National Bureau of Economic Research.

Hwang, M. and S. Park (2016). The Impact of Walmart Supercenter Conversion on Consumer Shopping Behavior. *Management Science*, 13.

Kaplan, G. and G. Menzio (2016, May). Shopping Externalities and Self-Fulfilling Unemployment Fluctuations. *Journal of Political Economy 124*(3), 771–825.

Kwon, S. Y., Y. Ma, and K. Zimmermann (2021). 100 Years of Rising Corporate Concentration. *SSRN Electronic Journal*.

Leung, J. H. (2021, October). Minimum Wage and Real Wage Inequality: Evidence from Pass-Through to Retail Prices. *The Review of Economics and Statistics 103*(4), 754–769.

Leung, J. H. and Z. Li (2022). Big-box Store Expansion and Consumer Welfare. Working Paper.

Melitz, M. J. and S. Polanec (2015, June). Dynamic Olley-Pakes productivity decomposition with entry and exit. *The RAND Journal of Economics 46*(2), 362–375.

Messinger, P. R. and C. Narasimhan (1997). A Model of Retail Formats Based on Consumers' Economizing on Shopping Time. *Marketing Science 16*(1), 1–23. Publisher: INFORMS.

Neiman, B. and J. Vavra (2021). The Rise of Niche Consumption. Working Paper.

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. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/iere.12343.

Radaelli, P. and M. Zenga (2002, May). A "New" Two Term Subtractive Decomposition of Herfindahl Concentration Measure. Volume Sessioni Spontanee, pp. 525–528. Società Italiana di Statistica.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2021, May). Diverging Trends in National and Local Concentration. *NBER Macroeconomics Annual 35*, 115–150. Publisher: The University of Chicago Press.

Smith, D. and S. Ocampo (2021). The Evolution of U.S. Retail Concentration. Working Paper.

Syverson, C. (2019, August). Macroeconomics and Market Power: Context, Implications, and Open Questions. *Journal of Economic Perspectives 33*(3), 23–43.

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), 2308–51.

Tables

Table 1: Effects of Entry: Number of Stores Within Own 5-digit Zip Code

	(1)	(2)	(3)	(4)
VARIABLES	Household Retail Concentration			
NumSup	0.00812***	0.00799***		
	(0.00254)	(0.00237)		
NumClubs			-0.0121***	-0.0129***
			(0.00281)	(0.00277)
Observations	134,495	1,837,668	350,865	2,468,557
R-squared	0.681	0.698	0.682	0.686
Prob > F	0.001	0.001	0.000	0.000
Household-Quarter FE	X	X	X	X
Year-Quarter FE	X	X	X	X
Number of units	9371	125817	21749	152934
Number of clusters	9371	125817	21749	152934
BOTE	0.1688	0.019	-0.0766	-0.014

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Columns (1) and (3) include households living in a 5-digit zip code with store entry during the period. Columns (2) and (4) include all households. NumSup refers to the number of supercenters and NumClubs refers to the number of club stores. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table 2: Effect of Variety and Prices on Household Retail Concentration

	(1)	(2)	(3)	(4)	(5)	(6)
Data	HMS				RMS	
VARIABLES	Household HHI					
Variety: UPCs	0.0210***		0.0206***		0.0913***	
	(0.00485)		(0.00438)		(0.0177)	
Variety: Modules		0.0992***		0.0897***		0.0619
		(0.0219)		(0.0188)		(0.0625)
RPI (County)	-0.136***	-0.108***				
	(0.0292)	(0.0306)				
RPI (US)			-0.227***	-0.215***		
			(0.0302)	(0.0308)		
Price Index					0.0184	-0.0205
					(0.0330)	(0.0347)
Observations	3491193	3491193	3605864	3605864	974443	974443
R-squared	0.663	0.661	0.662	0.660	0.696	0.696
Prob > F	0.000	0.000	0.000	0.000	0.000	0.296
Number of units	185449	185449	190795	190795	82135	82135
Number of clusters	185449	185449	190795	190795	82135	82135
First stage F-stat	400.448	72.449	428.909	72.872	1624.059	409.923
BOTE: All	0.027	0.034	0.025	0.030	0.160	0.002
BOTE: Variety	0.038	0.043	0.040	0.044	0.159	0.004
BOTE: Prices	-0.012	-0.009	-0.015	-0.014	0.001	-0.002

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety: UPCs refers to the number of UPCs per store, Variety: Modules refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 4.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table 3: Effect of shopping trips on household HHI, IV with household characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IV	None	None: Controls	Age	Size		Employment	
VARIABLES			Household HHI				
Log Trips	-0.129***	-0.201***	-0.204***	-0.203***	-0.198***	-0.218***	-0.222***
	(0.000364)	(0.000613)	(0.00291)	(0.0141)	(0.0145)	(0.0134)	(0.0137)
Observations	3668938	3622992	2312242	3622992	3622926	3622992	3622926
R-squared	0.712	0.724	0.384	0.724	0.739	0.724	0.739
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	194524	193527	2650	193527	193526	193527	193526
Number of clusters	194524	193527	136238	193527	193526	193527	193526
First stage F-stat			6225.057	102.427	98.491	116.170	109.195
BOTE: Trips	0.613	0.957	0.941	0.963	0.939	1.036	1.055
County-time FE			X		X		X

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Controls are control variables for shopping needs as in AH, which include the log of the quantity index derived from the RPI as well as the log number of UPCs and product groups purchased per period. BOTE refers to a back-of-the-envelope estimate of how much each independent variable (log trips) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table 4: Effect of shopping trips on household HHI, IVs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IVs		Real wage			Unemployment Rate			Both	
IV group	Age	Education	Both	Age	Education	Both	Age	Education	Both
VARIABLES					Household HHI				
Log Trips	-0.434***	-0.181***	-0.331***	-0.170***	-0.336***	-0.243***	-0.298***	-0.282***	-0.291***
	(0.0773)	(0.0659)	(0.0498)	(0.0420)	(0.0699)	(0.0373)	(0.0361)	(0.0563)	(0.0306)
Observations	2285235	2305700	2278148	2285235	2305700	2278148	2285235	2305700	2278148
R-squared	0.675	0.737	0.720	0.737	0.717	0.737	0.728	0.730	0.730
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128280	128764	128038	128280	128764	128038	128280	128764	128038
Number of clusters	128280	128764	128038	128280	128764	128038	128280	128764	128038
First stage F-stat	22.761	22.522	23.011	54.345	18.622	38.693	42.073	15.757	30.737
BOTE: Trips	2.182	0.839	1.674	0.814	1.607	1.191	1.541	1.341	1.535
BOTE: IVs	0.097	0.008	0.083	0.023	0.035	0.061	0.111	0.037	0.143

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. We construct proxies for labor demand shocks at various levels of aggregation across household characteristics in the CPS MORG, such as age and education cells or both across the US, which we denote as an IV group. These IV's include real wages, unemployment rates, or both, which we denote as IVs. We allow these variables to differ by gender. We match these proxies to households in the HMS using the mode of their characteristics over the sample period to avoid using variation induced by shifts in characteristics over time. BOTE: Trips refers to a back-of-the-envelope estimate of how much each independent variable (log trips) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period. To calculate the BOTEs resulting from changes in the IV only (BOTE: IVs), we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively.

Table 5: Comparative Statics: Household HHI and Welfare

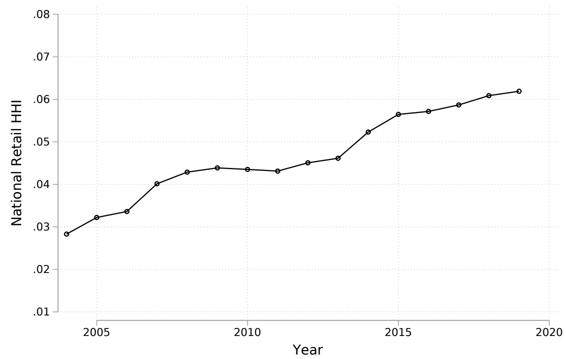
Event	Model Parameter	HHI	Consumer Welfare
Demand			
Wage rises	$w \uparrow$	$\uparrow$	$\uparrow$
Leisure technology rises	$\theta \uparrow$	$\uparrow$	$\uparrow$
Trip cost rises	$t \uparrow$	$\uparrow$	$\downarrow$
Supply			
A closer store	$\mu \downarrow$	$\downarrow$	$\uparrow$
Entry of a supercenter	$\underline{P^F} \downarrow$	$\uparrow^{*}$	$\uparrow$
Prices $\downarrow$ or variety $\uparrow$	$P^F \downarrow$	$\uparrow^{**}$	$\uparrow$

Notes: This table summarizes some comparative statics from our model, both on the supply and demand side, and their implications for household HHI and welfare. * denotes the fact that HHI increases when there is a sufficient decrease in P^{F*}. ** denotes the fact that HHI increases when the price index decrease favors firms with lower initial price indices as in equation (21).

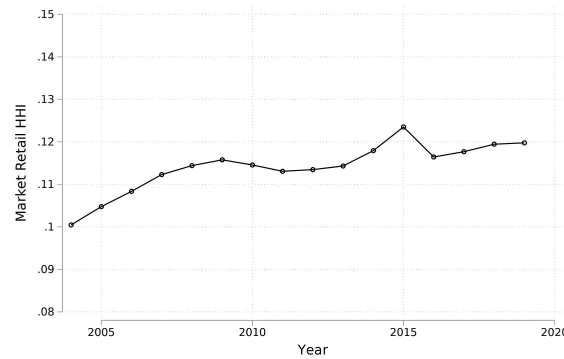
Figures

Figure 1: Retail Concentration Over Time

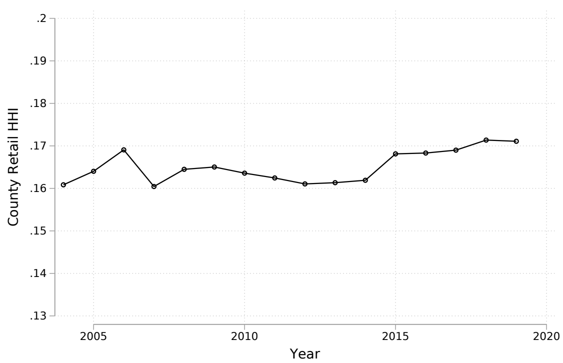
(a) National



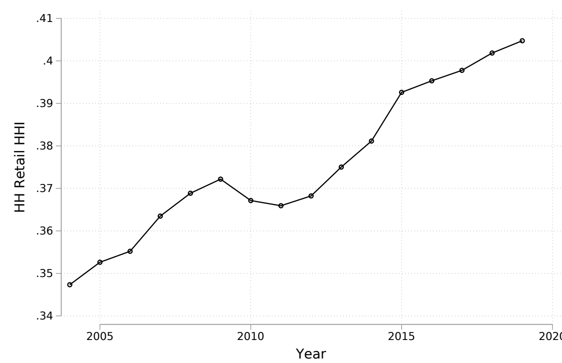
(b) Market



(c) County



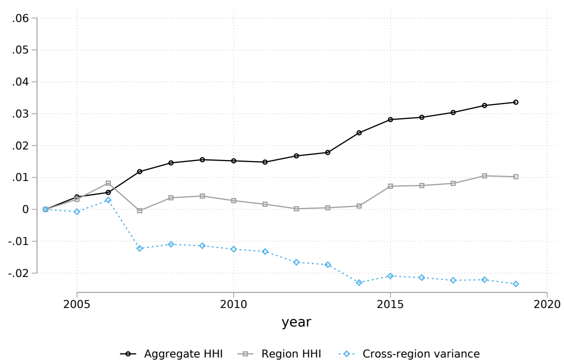
(d) Household



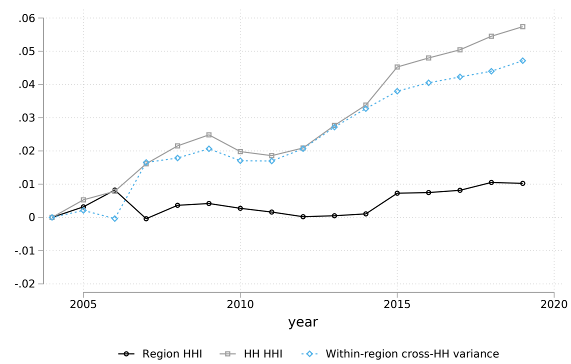
Notes: This figure plots the revenue-weighted average of retail concentration at the national, Nielsen Scantrack market, county, and household level respectively.

Figure 2: Decomposing Changes in Retail Concentration

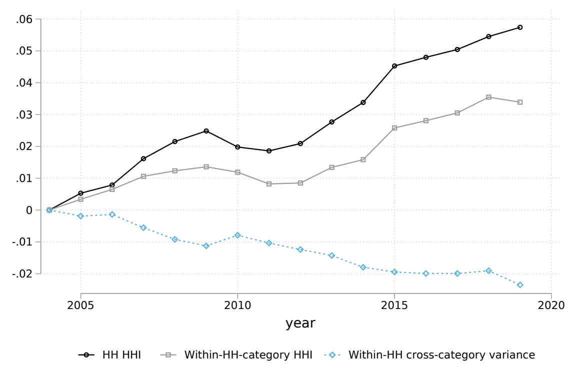
(a) National and County



(b) County and Household



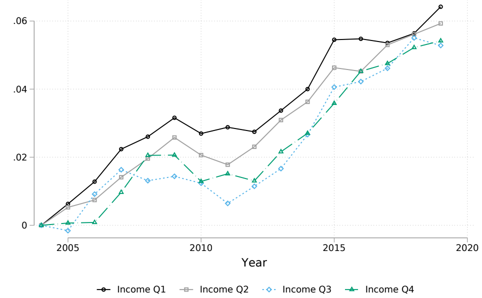
(c) Household and Category



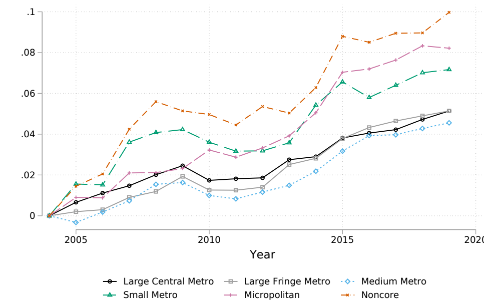
Notes: This figure shows that HHI can be decomposed at each level following Radaelli and Zenga (2002). For each term, we plot the yearly change over the sample period.

Figure 3: Changes in Household Retail Concentration by Demographic Group

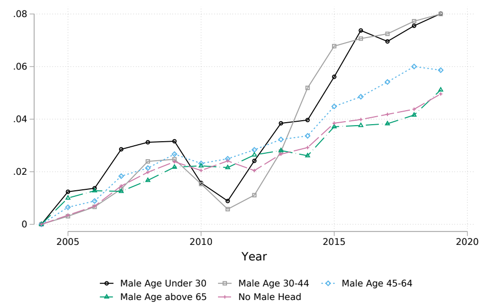
(a) Household Income



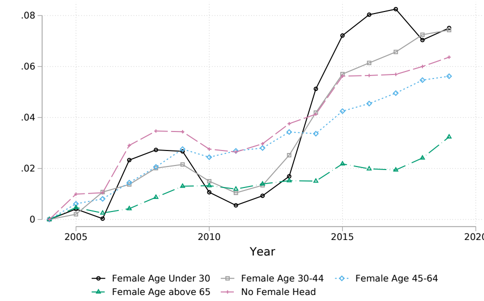
(b) Urban/Rural Counties



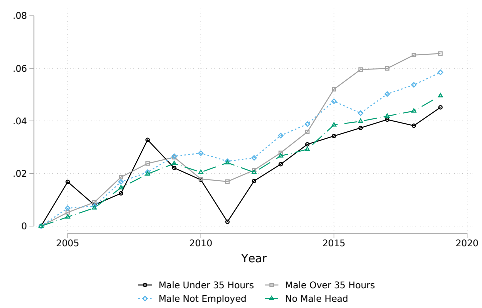
(c) Male Head Age



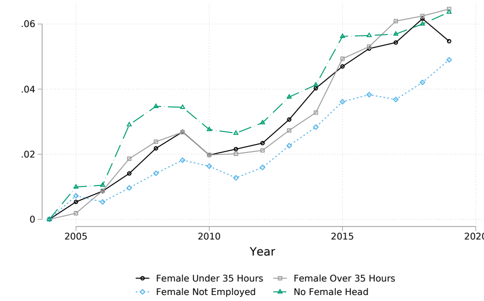
(d) Female Head Age



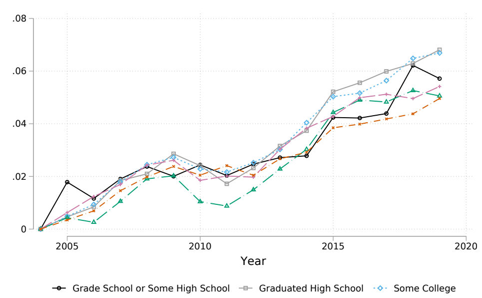
(e) Male Head Employment



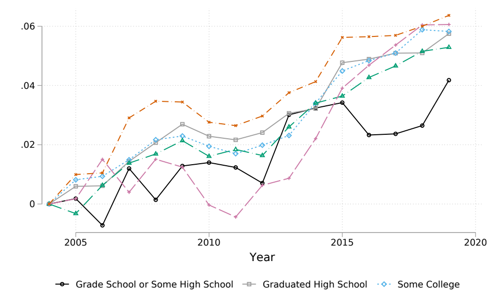
(f) Female Head Employment



(g) Male Head Education



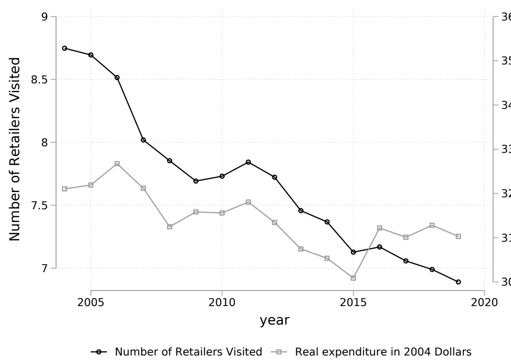
(h) Female Head Education



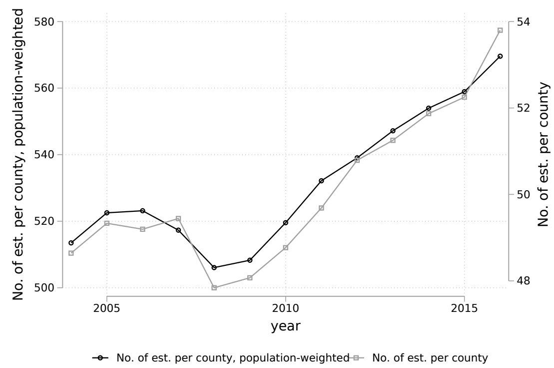
Notes: This figure plots the changes in revenue-weighted average of retail concentration at the household level for different demographic groups. Panel (a) classifies households into four household income quartiles after adjusting for household size, with Q1 being the lowest income quartile and Q4 the highest income quartile. Panel (b) classifies households by the county they reside in and groups them based on the urban/rural classification of the National Center for Health Statistics (NCHS). Panel (c) and (d) classifies households by their male and female head's age. Panel (e) and (f) classifies households by their male and female head's hours of work. Panel (g) and (h) classifies households by their male and female head's level of education.

Figure 4: Retail Chains Visited, Number of Drug, Grocery, and Mass Merchandise Establishments per County, and Shopping Trips and Days

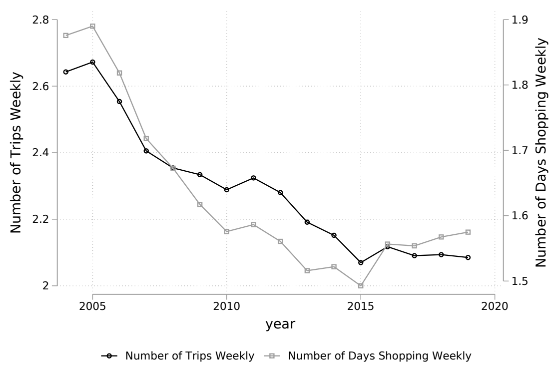
(a) Retail Chains Visited



(b) Establishments per County

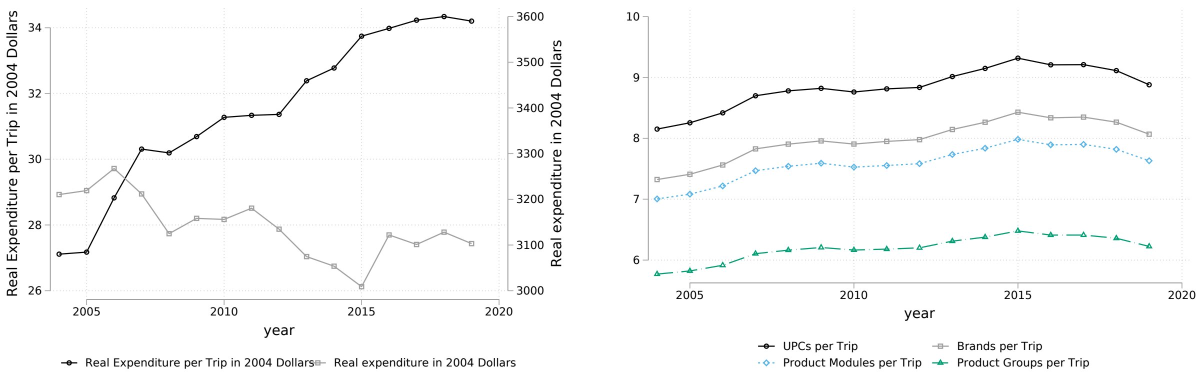


(c) Shopping Trips and Days



Notes: Figure 4a plots the number of retail chains visited per quarter per household, aggregated first to the yearly level with a simple average and then averaged across households weighted by sampling weights. We deflate the household expenditures using the chained food-at-home price index from the BLS. We obtain the total number of drug, grocery, and mass merchandise establishments per county from the County Business Patterns data from the Economic Census. We plot these numbers only up to 2016 due to changes in the reporting thresholds since 2017.

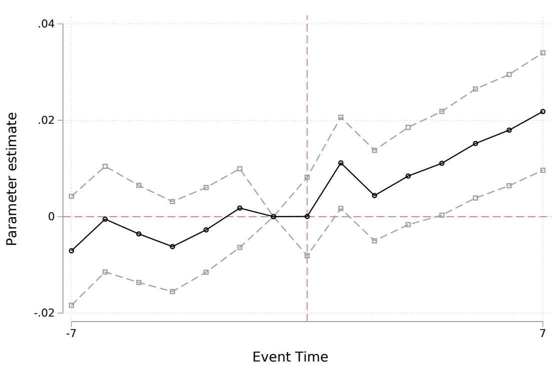
Figure 5: Expenditure per Trip and Varieties per Trip



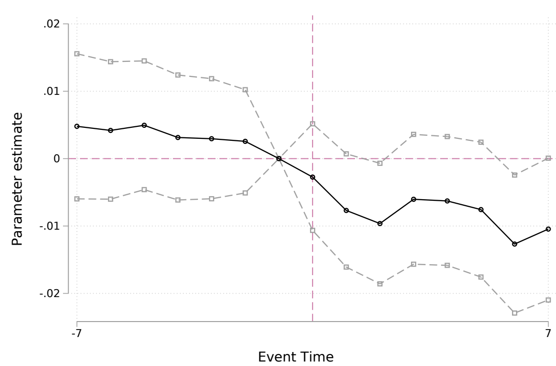
Notes: In Panel (a), this figure plots the average of real expenditure per shopping trip and real expenditure per year in 2004 dollars, deflated by the chained CPI and weighted by sampling weights. In Panel (b), this figure plots the UPCs, brands, product modules, and product groups per trip as defined by the HMS.

Figure 6: Event Study Graph for Store Entry

(a) Supercenter Entry



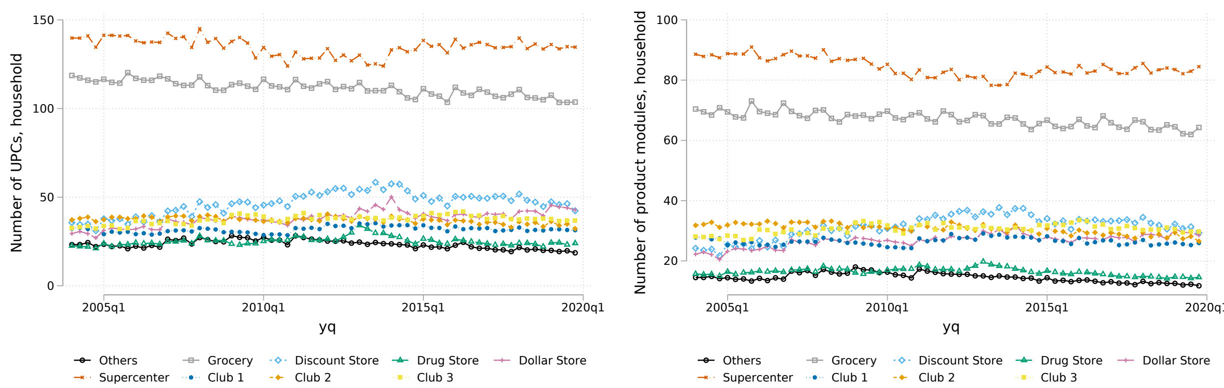
(b) Club Entry



Notes: This figure plots the sum of estimated coefficients for each period, along with the 95% confidence intervals, from regressions using a distributed lag model, where household retail HHI is regressed on the number of supercenters and club stores respectively. Household and period fixed effects are included.

Figure 7: Product Assortment in Big-box Stores

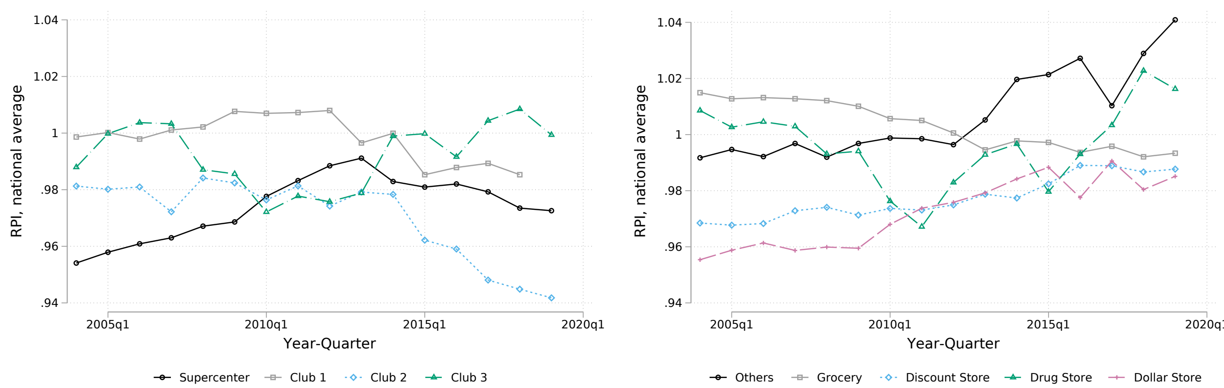
(a) UPCs (b) Product Modules



Notes: This figure plots the sample-weighted average number of UPCs and product modules households buy in each type of store per quarter.

Figure 8: RPI in Big-box Stores and Other Channel Types

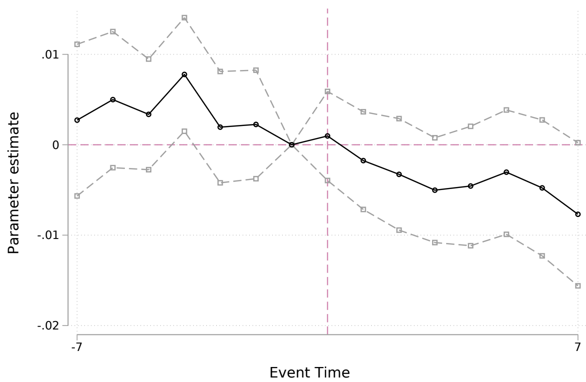
(a) Big-box Stores (b) Other Channel Types



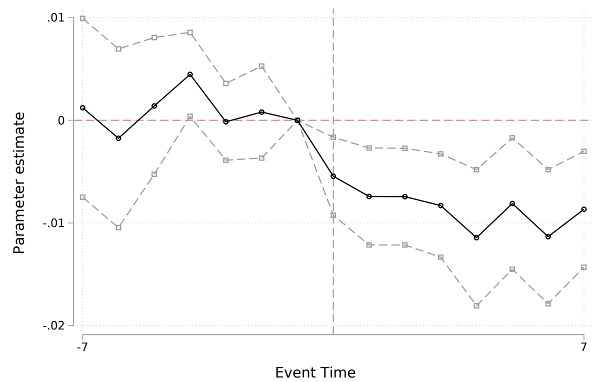
Notes: This figure plots the relative price index (RPI) of each retailer or channel type following AH. To construct a retailer RPI, we calculate the ratio between total expenditure for each good and the counterfactual expenditure of each good at its average price in the reference region. We then take the weighted average across goods and counties to calculate a national RPI for each retailer that uses national averages as reference prices.

Figure 9: Effects of non-employment, female heads

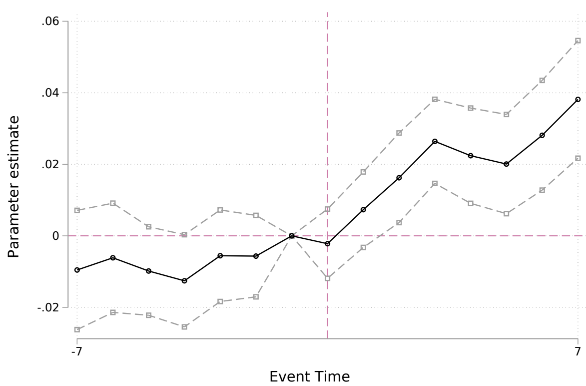
(a) Household HHI, OLS



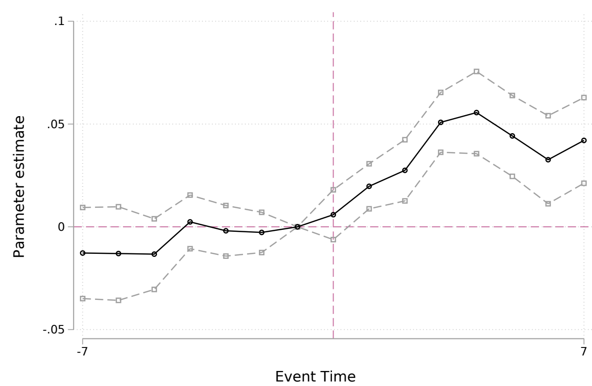
(b) Household HHI, DCDH



(c) Log trips, OLS



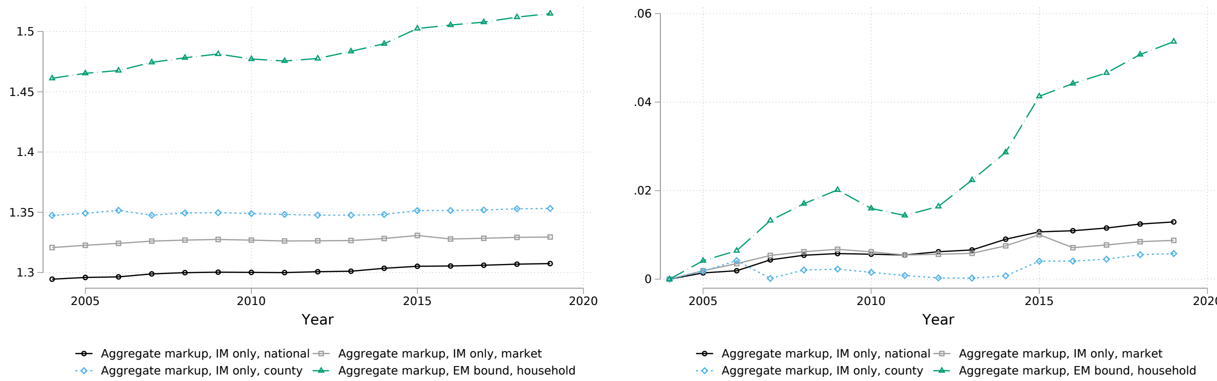
(d) Log trips, DCDH



Notes: This figure plots the sum of estimated coefficients for each period, along with the 95% confidence intervals, from regressions using a distributed lag model, where household retail HHI is regressed on an indicator for whether the household is not working. The sample focuses on prime-aged household heads from age 25-54. Household and period fixed effects are included. We use both OLS and methods robust to heterogeneous treatment effects following de Chaisemartin and D'Haultfœuille (2020).

Figure 10: Aggregate markups, 2004-2019

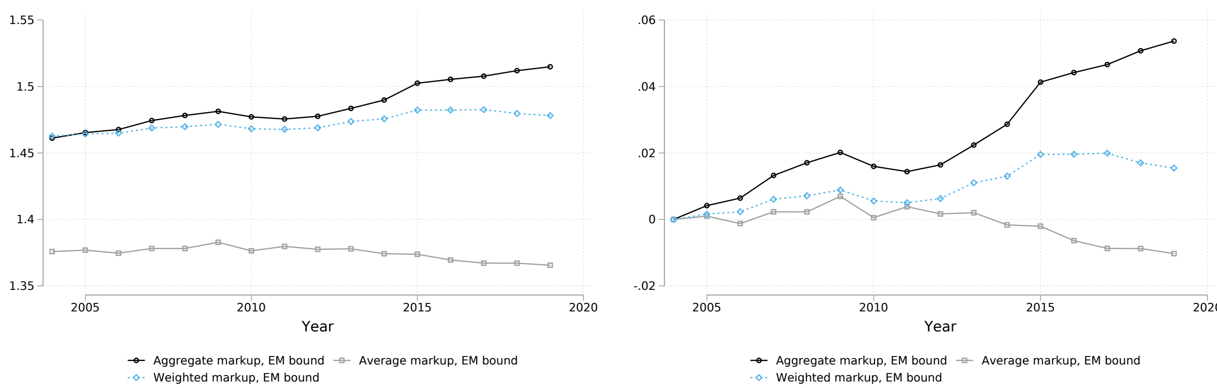
(a) Levels (b) Changes since 2004



Notes: This figure plots the trends in aggregate markup from 2004-2019 when calculated using different levels of aggregation, namely national, market, county, and household respectively. For the first three levels, we only calculate using the intensive margin (IM) term. For the household level, we can calculate using both intensive and extensive margin (EM) terms, using the upper bound for the EM.

Figure 11: Aggregate, average, and weighted markups, 2004-2019

(a) Levels (b) Changes since 2004



Notes: This figure plots the trends in firm markups from 2004-2019 when aggregated using different weights, assuming Bertrand competition, both in levels and the magnitude of changes since 2004. The default aggregate markup uses firm market shares over time. Average markup uses a simple average for firms that were present in the entire sample period. Weighted markup uses a weighted average, where weights are fixed to the aggregate market share across the entire sample period, for firms that were present in the entire sample period.

For Online Publication: Appendix

A Measuring Retail Concentration

We use the Herfindahl-Hirschman Index (HHI) as our primary measure of retail concentration at the household level and different geographical levels.

Household retail concentration is constructed in three steps. First, for each household i and retailer j, we calculate the total expenditure S_{ij}^t in period t. Second, the associated market share is

$$m_{ij}^t = \frac{S_{ij}^t}{\sum_j S_{ij}^t}.$$

Finally, household retail concentration is the HHI of the above market share:

$$H_i^t = \sum_j (m_{ij}^t)^2.$$

We further calculate the regional average household retail concentration, defined as the weighed average of household retail concentration in a given region. Let $r(i,t)$ denote the region where household i lives in period t. The set of households who lives in region r in period t is $I_{rt} = \{i : r(i,t) = r\}$. Regional average household retail concentration can then be calculated from

$$\bar{H}_r^t = \sum_{i \in I_{rt}} \alpha_{ir}^t H_i^t.$$

The household weight α_{ir}^t depends on all the households' sampling weight (projection factor) w_{ir}^t and total expenditure S_{ij}^t:

$$\alpha_{ir}^t = \frac{\sum_j (w_{ir}^t S_{ij}^t)}{\sum_{i:r(i,t)=r} \sum_j (w_{ir}^t S_{ij}^t)}.$$

We use a similar measure for retail concentration at different geographical units, such as counties, Scantrack markets as defined by Nielsen, and nationally. Following the above definition, the market share for retailer j at time t in region r is

$$m_{jr}^t = \frac{\sum_{i \in I_{rt}} (w_{ir}^t S_{ij}^t)}{\sum_j \sum_{i \in I_{rt}} (w_{ir}^t S_{ij}^t)}.$$

The HHI of region r in period t becomes

$$H_r^t = \sum_j (m_{jr}^t)^2.$$

We call it the aggregate retail concentration (or region HHI) of region r in period t, which is a measure of regional retail concentration.

An alternative measure of retail concentration is the total share of expenditure on the top n retail chains, which is generally known as concentration ratios C_n. Since results are qualitatively identical using this measure, we focus on the HHI in our main analysis.

A.1 RZ Decomposition

Moreover, we follow Radaelli and Zenga (2002) (RZ) to define the cross-household variance of retail concentration:

$$V_r^t = \sum_{i:r(i,t)=r} \alpha_{ir}^t \left(\sum_j (m_{ij}^t - m_{jr}^t)^2 \right).$$

We can formally relate region average household concentration and region aggregate concentration with the decomposition:

$$H_r^t = \bar{H}_r^t - V_r^t.$$

This decomposition suggests that if all the households' consumption patterns in each retail chain are the same, then $\bar{V}_r^t = 0$ and the region average household retail concentration is a perfect reflection of the region aggregate concentration in the region.

Therefore, beginning at the national level, we can decompose changes in national aggregate HHI $\Delta H = H^t - H^{t-n}$ into changes in the national average region HHI minus the cross-region variance:

$$\Delta H = \Delta \sum_r \alpha_r H_r - \Delta \sum_r \alpha_r V_r$$
$$\Delta\text{National HHI} = \Delta\text{Region HHI} - \Delta\text{Cross-region variance.}$$

At the regional level, we can decompose changes in the region aggregate HHI into changes in the region average household HHI minus the cross-household variance:

$$\Delta H_r = \Delta \sum_r \sum_i \alpha_{ir} H_{ir} - \Delta \sum_r \sum_i \alpha_{ir} V_{ir}.$$
$$\Delta\text{Region HHI} = \Delta\text{Household HHI} - \Delta\text{Cross-household variance.}$$

At the household level, we can further classify expenditures at the household-product-category level S_{ic}. We then decompose changes in the household aggregate HHI into changes in the household average category HHI minus the cross-category variance:

$$\Delta H_{ir} = \Delta \sum_r \sum_i \sum_c \alpha_{irc} H_{irc} - \Delta \sum_r \sum_i \sum_c \alpha_{irc} V_{irc}$$
$$\Delta\text{Household HHI} = \Delta\text{Household-category HHI} - \Delta\text{Cross-category variance.}$$

Combining all four levels, changes in national aggregate HHI can be formally decomposed into changes in the following four terms: Household-category HHI H_{irc}, household cross-category variance V_{irc}, regional cross-household variance V_{ir}, and national cross-region variance V_r. We write out the entire decomposition as follows:

$$\begin{aligned}
\Delta H &= \Delta \sum_r \sum_i \sum_c \alpha_{irc} H_{irc} - \Delta \sum_r \sum_i \sum_c \alpha_{irc} V_{irc} \\
&\quad - \Delta \sum_r \sum_i \alpha_{ir} V_{ir} - \Delta \sum_r \alpha_r V_r \\
\Delta\text{National HHI} &= \Delta\text{Household-category HHI} - \Delta\text{Cross-category variance} \\
&\quad - \Delta\text{Cross-household variance} - \Delta\text{Cross-region variance},
\end{aligned}$$

where

$$H_{irc} = \sum_j (m_{irjc})^2, \;\; V_{irc} = \sum_j (m_{irjc} - m_{irj})^2, \;\; \alpha_{irc} = \frac{w_{ir} S_{irc}}{\sum_i \sum_r \sum_c w_{ir} S_{irc}}.$$

To investigate whether composition changes in our dataset, such as the entry and exit of regions or households, are driving our results, we also use a Dynamic Olley-Pakes decomposition (Melitz and Polanec 2015) on top of the RZ decomposition, as shown below:

$$\begin{aligned}
&\Delta \sum_r \sum_i \sum_c \alpha_{irc} H_{irc} \\
&= (H_{S2} - H_{S1}) + \alpha_{E2} (H_{E2} - H_{S2}) + \alpha_{X1} (H_{S1} - H_{X1}) \\
&= \Delta \bar{H}_S + \Delta cov_S + \alpha_{E2} (H_{E2} - H_{S2}) + \alpha_{X1} (H_{S1} - H_{X1}) \\
&\alpha_{Gt} = \sum_{i \in G} \alpha_{it}, \;\; H_{Gt} = \sum_{i \in G} \frac{\alpha_{it}}{\alpha_{Gt}} H_{it}
\end{aligned}$$

Applying this decomposition to each of the four terms decomposing the national aggregate HHI in Appendix Table F3, we find that within-survivor growth is driving most of the variation.

A.2 RZ Decomposition by Firm

Moreover, we can show that the aggregate HHI, average region HHI, and cross-region variance can be decomposed into contributions by various groups of firms (e.g. group $\mathcal{J}_1$ and $\mathcal{J}_2$) as follows:

$$
\begin{aligned}
H &= \sum_{j\in\mathcal{J}_1}(m_j)^2 + \sum_{j\in\mathcal{J}_2}(m_j)^2 \\
H &= \sum_r \alpha_r H_r - \sum_r \alpha_r V_r \\
&= \sum_r \alpha_r \sum_j (m_{jr})^2 - \sum_r \alpha_r \sum_j (m_{jr} - m_j)^2 \\
&= \sum_r \alpha_r \left(\sum_{j\in\mathcal{J}_1}(m_{jr})^2 + \sum_{j\in\mathcal{J}_2}(m_{jr})^2 \right) - \sum_r \alpha_r \left(\sum_{j\in\mathcal{J}_1}(m_{jr} - m_j)^2 + \sum_{j\in\mathcal{J}_2}(m_{jr} - m_j)^2 \right) \\
&= \sum_{j\in\mathcal{J}_1}\sum_r \alpha_r (m_{jr})^2 + \sum_{j\in\mathcal{J}_2}\sum_r \alpha_r (m_{jr})^2 - \sum_{j\in\mathcal{J}_1}\sum_r \alpha_r (m_{jr} - m_j)^2 - \sum_{j\in\mathcal{J}_2}\sum_r \alpha_r (m_{jr} - m_j)^2
\end{aligned}
$$

B Robustness of Household HHI Trends

In this section, we show that the rise of household HHI is robust to potential changes to the sample of households in the HMS. First, we show that our results are robust to adding various controls in Appendix Figure G8. In Figure G8a, we regress household HHI on various control variables and plot the year fixed effects for each specification. We control for the number of years the household has been in the sample, total expenditures, household fixed effects, as well as household and regional demographics. All of these specifications continue to show a rising trend for household HHI. In Figure G8b, we show that our sample captures an average of real expenditures that is roughly constant over the sample period.

Second, we show that the flat trend in real expenditures is consistent with a flat trend for the same set of products in the Consumption Expenditure Survey (CEX). In Figure G9, we show the average annual expenditure of food-at-home and matched non-durable goods captured by Nielsen, as classified by Coibion et al. (2021), in the Consumption Expenditure Survey (CEX) Interview Survey (IS) from 2004-2015 using the latest available NBER CEX data. We show these trends with and without controls for various household demographics. Household expenditures in the non-magnet sample, with an annual average of roughly \$3200, covers over 60% of matched non-durable goods in the CEX, which has an annual average of roughly \$5000.

Third, our main results use only products with barcodes. Nielsen also captures products without standard barcodes using their own classification system. This data, known as magnet data, includes items such as fruits, vegetables, meats, and in-store baked goods. However, only a subset of households regularly report magnet product purchases starting from 2007. These magnet households have been increasing substantially over the sample period, with their proportion increasing from about 35% in 2007 to 87% in 2019. Using only magnet households, we show that the average household HHI has been decreasing for magnet products, although this trend is not very robust across specifications. This could be driven by large changes in the magnet sample, and we see large fluctuations in average real expenditures over the sample period. We show in Figure G11 that average household HHI continues to show an upward trend using all products for magnet households, although the trend is slightly attenuated relative to the barcode-only sample. Since the trend in average real expenditures again shows a lot of fluctuation, we focus on using only the non-magnet sample for our main results.

C Other Hypotheses

What are some other hypotheses that can explain the rise in household concentration? In Appendix Section D, we discuss the rise of online shopping as a related hypothesis, which may also induce households to engage in more one-stop shopping. We find that concentration trends barely change when considering only offline retailers, given that the share of online shopping remains small in our sample period. This is consistent with Hortaçsu and Syverson (2015), who show evidence that online share of retail sales remains very small in the product categories we consider. They argue that although online retail will surely continue to be a force shaping the sector going forward and may yet emerge as the dominant mode of commerce in the retail sector in the US, its time for supremacy has not yet arrived. We also consider a few empirical strategies and find that the effect of online shopping is again negligible.

We also investigate changes in the cost of travel as a potential hypothesis. Following Gelman et al. (2016), who argue that gas prices during 2014-2015 were driven by unanticipated, permanent, and exogenous supply shocks, we estimate the effect of the number of shopping trips on household HHI using state- or city-level gas prices from the Energy Information Administration (EIA) and Bureau of Labor Statistics (BLS) as IVs. We focus on the period of 2010-2016 to avoid using recessionary shocks that may also be driven by the demand side. Our identifying assumption is that the elasticity of gas prices to supply shocks, i.e. demand elasticity for gas, is plausibly exogenous to unobservables affecting household HHI. Note that we have controlled for shopping needs to attempt to account for income effects. We show our results in Appendix Table F17. We wish to interpret these results with caution, since the first-stage F-statistic is quite weak, ranging from 4-8. The BOTE implies that the rise in gas prices from 2004-2019 could explain up to 12% of the rise in household HHI, but since gas prices actually fell from 2010-2016 as shown in Appendix Figure G21, changes in gas prices cannot explain the rise of household HHI during this period. Nonetheless, these results provide suggestive evidence that a rise in the cost of travel, in this case gas prices, does lower shopping time and raise household HHI.

In addition to the cost of travel, factors such as a rise in the preference for leisure and a rise in leisure technology can also increase the opportunity cost of shopping time. Our theoretical results in Section 5 show that these factors can indeed raise both household concentration and cross-household variance. Using the ATUS sample from Aguiar et al. (2021), we find that the drop in weekly shopping time from 2004-2007 to 2014-2017 is smaller for men relative to women, which stands in contrast with the rise of recreational computing being the largest for young men. Shopping time fell by about 18% for women aged 21-30 from a base of 6.2 hours to 5.1 hours, and 15% for women aged 31-55 from a base of about 6.5 hours to 5.5 hours. Men aged 21-30 experienced a fall of about 10% from a base of 4 hours to 3.6 hours, while men aged 31-55 experienced the smallest drop of 5% from a base of 4.1 hours to 3.9 hours. Table 1 of Aguiar et al. (2021) shows that the rise in leisure was largest for young men (2.3 hours) but also significant for women (1.6 and 1.3 hours), and smallest for older men (0.6 hours). Therefore, the rise in recreational computing technology seems insufficient to explain the rise in household concentration for both younger and older women as shown in Figure 3d. Other forms of leisure technology that impact women in both of these age groups could be a contributing factor.[36]

We also consider the possibility of whether tastes are converging within households, which may be the case if there is increasing positive assortative matching on tastes in household formation. Appendix Figure G7 shows that the increase in household HHI holds across all household sizes and is slightly larger for households with more members. Given that single-member households experienced a substantial increase in household HHI, we believe that converging tastes within households is unlikely to be a key explanation for rising household concentration.

Our results on cross-household variance are also consistent with increasing divergence in tastes between households. However, our theoretical results in Section 5.4 imply that diverging tastes are not necessary to generate increases in cross-household variance. As long as there is heterogeneity in shopping costs across households, even if the degree of heterogeneity remains unchanged over time, a rise in the opportunity cost of shopping time can generate rising cross-household variance because households will polarize to their preferred retailers and the overlap in choice sets will decrease. We leave an in-depth exploration of diverging tastes for future research.

The rise in household concentration and cross-household variance can also be consistent with increasing consumer loyalty. For example, club stores require customers to become members in order to shop there, which may encourage households to only shop at retailers at which they are members. Our results show that a club store entry actually lowers household HHI, but this is insufficient to rule out loyalty as a contributing factor since club stores could lower household HHI for other reasons that we discussed previously such as lower product variety, potentially offsetting some of the loyalty effect. Therefore, we leave loyalty as an explanation for further exploration in future research.

We also investigate the differences across regions based on urban-rural classifications interacted with household income in Appendix Figure G7. We find that differences between income groups remain similar even after conditioning on urban-rural classification. Given the magnitude of increases are similar between income groups, we believe that changes in income inequality and polarization are unlikely to be key drivers of rising household concentration.

D Online Shopping

Can the rise of online shopping contribute to the trends we observe? We find that concentration trends barely change when considering only offline retailers in Appendix Figure G22, given that the share of online shopping remains small even as it is increasing. This is consistent with Hortaçsu and Syverson (2015), who show evidence that online share of retail sales remains very small in the product categories we consider. They argue that although online retail will surely continue to be a force shaping the sector going forward and may yet emerge as the dominant mode of commerce in the retail sector in the US, its time for supremacy has not yet arrived.

[36]While we do not have a simple empirical strategy to quantify the size of these effects, one possibility could be to estimate a leisure demand system following Aguiar et al. (2021).

In Appendix Figure G23, we further show the share of online shopping in the product departments in our data. While the shares are increasing in nearly every product department, they all remain below 5%. Appendix Figure G24 shows the household concentration trends are very similar with and without online retailers.

We investigate the impact of online shopping shares on household HHI in Appendix Table F18. We first regress the household HHI with offline retailers, household HHI, the number of offline shopping trips, and the number of offline retailers visited on the online shopping share, with household and time fixed effects. Given that online shopping share is likely to be endogenous, we caution against interpreting these estimates as causal and use them as a rough estimate of the potential impacts of online shopping. In fact, if some unobserved factors lead to both higher household concentration and higher share of online shopping (e.g. higher trip costs), the effect of online shopping will be overestimated. We can then treat the estimates as upper bounds. We find that since the rise in online shares remain small, the BOTEs are at most 1-2%. Given that endogeneity is unlikely to substantially downward bias our estimates, we view these results as suggestive evidence that online shopping is not a main driver of our trends.

To assuage potential endogeneity concerns, we use state-level variation in Amazon taxes to estimate the impact of online shopping on household HHI and the number of trips in Appendix Table F19. We find that the introduction of an Amazon tax has negligible effects on household HHI and the number of trips.

E Derivations

E.1 Demand

Consumers have a Cobb-Douglas utility function and derive utility from (1) quantities of a composite consumption good $X(\mathcal{V})$, (2) the total variety of these goods $\mathcal{V}$, and (3) leisure $L(\mathcal{V})$, with a preference for leisure of ρ as shown in equation (29). They maximize utility by choosing $X(\mathcal{V})$, $\mathcal{V}$, $L(\mathcal{V})$, and labor supply h, subject to a time constraint and budget constraint. The amount of time they have is T, which they split between shopping time $\tau(\mathcal{V})$, leisure L, and labor supply h. Shopping time is the integral of the shopping cost for each variety $\mu(\nu)$ over all varieties $\mathcal{V}$. They have income Y from supplying labor h at wage w and non-labor income K, which is used to buy consumption goods with price $p(\nu)$ for each variety ν. We can combine the time and budget constraints into the full income constraint in equation (30).

$$\max_{X,\mathcal{V},L,h} U(X(\mathcal{V}), L(\mathcal{V})) = X(\mathcal{V})^{1-\rho} L(\mathcal{V})^{\rho} \tag{29}$$

$$s.t. \quad \tau(\mathcal{V}) + L + h = T, \text{ where } \tau(\mathcal{V}) = \int_{\nu\in\mathcal{V}} \mu(\nu) d\nu$$

$$s.t.\ Y = wh + K = \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu = Px(\mathcal{V})$$

$$s.t.\ wT + K = \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu + w(\tau(\mathcal{V}) + L) \tag{30}$$

Consumption good $X(\mathcal{V})$ contains two CES nests for firms and products as shown in equation (31). The first nest consists of firms ν, which are retailers in our context. The second nest contains products u, which enables consumers to choose a set of products $\mathcal{U}_f$ from firm f, with consumers choosing the set of multiproduct firms $\mathcal{V}$.[37]

$$X(\mathcal{V}) = \left(\int_{\nu \in \mathcal{V}} \left(\varphi^F(\nu) x^F(\nu) \right)^{\frac{\sigma^F - 1}{\sigma^F}} d\nu \right)^{\frac{\sigma^F}{\sigma^F - 1}}, \; x^F(\mathcal{U}_f) = \left(\int_{u \in \mathcal{U}_f} \left(\varphi^U(u) x^U(u) \right)^{\frac{\sigma^U - 1}{\sigma^U}} du \right)^{\frac{\sigma^U}{\sigma^U - 1}} \tag{31}$$

We solve for the demand equation for firms with the Lagrangian $\mathcal{L}$ and apply standard CES algebra as shown in Bronnenberg (2015).

$$\begin{aligned}
\mathcal{L} = U(X(\mathcal{V}), L(\mathcal{V})) + \lambda \left(wT + K - w\left(\tau(\mathcal{V}) + L\right) - \int_{\nu \in \mathcal{V}} p(\nu) x(\nu) d\nu \right) \\
\frac{\partial \mathcal{L}}{\partial x} = (1-\rho) \left(\int_{\nu \in \mathcal{V}} (\varphi(\nu) x(\nu))^{\frac{\sigma-1}{\sigma}} d\nu \right)^{\frac{\sigma}{\sigma-1}(1-\rho)-1} L(\mathcal{V})^{\rho} (\varphi(\nu))^{\frac{\sigma-1}{\sigma}} x(\nu)^{-\frac{1}{\sigma}} - \lambda p(\nu) = 0 \\
(1-\rho) X(\mathcal{V})^{-\rho+\frac{1}{\sigma}} L(\mathcal{V})^{\rho} \int_{\nu \in \mathcal{V}} (\varphi(\nu) x(\nu))^{\frac{\sigma-1}{\sigma}} d\nu = \lambda \int_{\nu \in \mathcal{V}} p(\nu) x(\nu) d\nu \\
\frac{(1-\rho) X(\mathcal{V})^{-\rho} L(\mathcal{V})^{\rho}}{P} = \lambda \quad \because Y = PX(\mathcal{V}) \\
\text{Standard CES FOC} \Rightarrow x(\nu) = A(\mathcal{V}) p(\nu)^{-\sigma} \varphi(\nu)^{\sigma-1} \\
A(\mathcal{V}) = Y P^{\sigma-1} \\
P = \left(\int_{\nu \in \mathcal{V}} \left(\frac{p(\nu)}{\varphi(\nu)} \right)^{1-\sigma} d\nu \right)^{\frac{1}{1-\sigma}}
\end{aligned}$$

Solving for the demand at the firm and product level following Hottman et al. (2016), we have the following equations:

[37]To maintain consistency with previous literature, we use ν and f interchangeably for firms.

$$
\begin{aligned}
x^F(\nu) &= A(\mathcal{V}) p^F(\nu)^{-\sigma^F} \varphi^F(\nu)^{\sigma^F-1} \\
A(\mathcal{V}) &= Y P^{\sigma-1},\ P = \left(\int_{\nu\in\mathcal{V}} \left(\frac{p^F(\nu)}{\varphi^F(\nu)} \right)^{1-\sigma^F} d\nu \right)^{\frac{1}{1-\sigma^F}} \\
x^U(u) &= \left(\varphi_f^F\right)^{\sigma^F-1} (\varphi_u^U)^{\sigma^U-1} Y P^{\sigma^F-1} (P_f^F)^{\sigma^U-\sigma^F} (P_u^U)^{-\sigma^U} \\
&= x^F(\nu)(\varphi_u^U)^{\sigma^U-1} \left(\frac{P_f^F}{P_u^U} \right)^{\sigma^U} \\
P_f^F &= \left(\int_{u\in\mathcal{U}_f} \left(\frac{P_u^U}{\varphi_u^U} \right)^{1-\sigma^U} \right)^{\frac{1}{1-\sigma^U}} \\
&= \left(\underbrace{N_f^U}_{\text{Scope}} \quad \underbrace{\frac{1}{N_f^U} \int_{u\in\mathcal{U}_f} \left(\frac{P_u^U}{\varphi_u^U} \right)^{1-\sigma^U}}_{\text{Average product taste-adjusted prices}} \right)^{\frac{1}{1-\sigma^U}}
\end{aligned}
$$

Similarly, using the first-order condition (FOC) for leisure, we have

$$
\frac{\partial \mathcal{L}}{\partial L} = \rho L(\mathcal{V})^{\rho-1} X(\mathcal{V})^{1-\rho} = \lambda w
$$

$$
L(\mathcal{V}) = \left(\frac{\lambda w}{\rho X(\mathcal{V})^{1-\rho}} \right)^{\frac{1}{\rho-1}} = \left(\frac{\lambda w}{\rho} \right)^{\frac{1}{\rho-1}} X(\mathcal{V}) \longrightarrow \frac{\rho L(\mathcal{V})^{\rho-1} X(\mathcal{V})^{1-\rho}}{w} = \lambda
$$

We also show the standard CES result that income equals the price index times quantity consumed.

$$
\begin{aligned}
\text{We want to show} \quad Y &= \int_{\nu\in\mathcal{V}} p(\nu)x(\nu)d\nu = PX(\mathcal{V}) \\
&= \left(\int_{\nu\in\mathcal{V}} \left(\frac{p(\nu)}{\varphi(\nu)}\right)^{1-\sigma} d\nu\right)^{\frac{1}{1-\sigma}} \left(\int_{\nu\in\mathcal{V}} (\varphi(\nu)x(\nu))^{\frac{\sigma-1}{\sigma}} d\nu\right)^{\frac{\sigma}{\sigma-1}} \\
\text{FOC: } \varphi(\nu)x(\nu) &= YP^{\sigma-1}\left(\frac{p(\nu)}{\varphi(\nu)}\right)^{-\sigma} \\
(\varphi(\nu)x(\nu))^{\frac{\sigma-1}{\sigma}} &= \left(YP^{\sigma-1}\left(\frac{p(\nu)}{\varphi(\nu)}\right)^{-\sigma}\right)^{\frac{\sigma-1}{\sigma}} \\
\int_{\nu\in\mathcal{V}} (\varphi(\nu)x(\nu))^{\frac{\sigma-1}{\sigma}} d\nu &= (YP^{\sigma-1})^{\frac{\sigma-1}{\sigma}} \int_{\nu\in\mathcal{V}} \left(\frac{p(\nu)}{\varphi(\nu)}\right)^{1-\sigma} d\nu \\
X(\mathcal{V}) &= YP^{\sigma-1}P^{(1-\sigma)\frac{\sigma}{\sigma-1}} = YP^{-1} \Rightarrow Y = PX(\mathcal{V})
\end{aligned}
$$

Given a Cobb-Douglas utility function for consumption and leisure, we combine these results with the CES utility to obtain the optimal leisure and consumption.

$$
\max_{c,\ell} U(c,\ell) = c^{1-\rho}\ell^{\rho} \ s.t. \left.\begin{aligned} p_c c &= wh + K \\ h + \ell + \tau &= T \end{aligned}\right\} p_c c + w\ell + w\tau = wT + K
$$

$$
\mathcal{L} = c^{1-\rho}\ell^{\rho} + \lambda(wT + K - w\ell - w\tau - p_c c)
$$

$$
\left.\begin{aligned} \tfrac{\partial \mathcal{L}}{\partial c} &= (1-\rho)c^{1-\rho-1}\ell^{\rho} - \lambda p_c = 0 \\ \tfrac{\partial \mathcal{L}}{\partial \ell} &= \rho\ell^{\rho-1}c^{1-\rho} - \lambda w = 0 \end{aligned}\right\} \lambda = \frac{(1-\rho)c^{-\rho}\ell^{\rho}}{p_c} = \frac{\rho\ell^{\rho-1}c^{1-\rho}}{w} \Rightarrow c = \ell^{\rho-\rho+1}\cdot\frac{1-\rho}{\rho}\frac{w}{p_c}
$$

$$
\begin{aligned}
c &= \frac{1-\rho}{\rho}\frac{w\ell}{p_c} \Leftrightarrow \ell = \frac{\rho}{1-\rho}\frac{c}{w}p_c \\
\frac{1-\rho}{\rho}w\ell + w\ell + w\tau &= wT + K \\
\frac{1}{\rho}w\ell &= wT - w\tau \\
\ell &= \frac{\rho\left(w(T-\tau)+K\right)}{w} \\
c &= \frac{(1-\rho)\left(w(T-\tau)+K\right)}{p_c} \\
\lambda &= (1-\rho)\left((1-\rho)\frac{(w(T-\tau)+K)}{p_c}\right)^{-\rho}\left(\rho\left(w(T-\tau)+K\right)\right)^{\rho}\frac{1}{p_c} \\
&= (1-\rho)^{1-\rho}\left(\frac{w}{p_c}\right)^{-\rho}\rho^{\rho}\frac{1}{p_c} = \rho^{\rho}(1-\rho)^{1-\rho}\frac{1}{\rho p_c^{1-\rho}w^{\rho}}
\end{aligned}
$$

$$
\begin{aligned}
\therefore L(\mathcal{V}) &= \frac{\rho\left(w(T-\tau(\mathcal{V}))+K\right)}{w} \\
X(\mathcal{V}) &= \frac{(1-\rho)\left(w(T-\tau(\mathcal{V}))+K\right)}{P} = \frac{Y}{P} = \frac{w(T-\tau)-wL+K}{P} \\
Y &= (1-\rho)\left(w(T-\tau(\mathcal{V}))+K\right) \\
&= w(T-\tau(\mathcal{V})-L(\mathcal{V}))+K
\end{aligned}
$$

To incorporate the possibility of changes in leisure technology in a simple manner, assume leisure can be scaled by a technology factor of θ as follows, which gives the following equation for consumption and leisure:

$$
\max_{c,\ell} u(c,\ell) = c^{1-\rho}(\ell^{\theta})^{\rho} \ s.t.\ p_c c + w\ell + w\tau = wT + K
$$

$$
\left.
\begin{aligned}
\tfrac{\partial \mathcal{L}}{\partial c} &= (1-\rho)c^{1-\rho-1}\left(\ell^{\theta}\right)^{\rho} - \lambda p_c = 0 \\
\tfrac{\partial \mathcal{L}}{\partial \ell} &= \theta\rho\ell^{\theta\rho-1}c^{1-\rho} - \lambda w = 0
\end{aligned}
\right\} \lambda = \frac{(1-\rho)c^{-\rho}\ell^{\theta\rho}}{p_c} = \frac{\theta\rho\ell^{\theta\rho-1}c^{1-\rho}}{w}
$$

$$
\begin{aligned}
\Rightarrow c &= \frac{1-\rho}{\theta\rho}\frac{w\ell}{p_c} \Rightarrow \frac{1-\rho}{\theta\rho}w\ell + w\ell + w\tau = wT + K \\
\ell &= \frac{\theta\rho}{1-\rho+\theta\rho}\frac{w(T-\tau)+K}{w} \\
c &= \frac{1-\rho}{1-\rho+\theta\rho}\frac{w(T-\tau)+K}{p_c}
\end{aligned}
$$

WLOG, assume $\nu \in \mathcal{V} = D$ and for brevity, suppress F and let $\sigma = \sigma^F$. We can derive the optimal variety using the indirect utility function by substituting the optimal $x(\nu)$ into the utility function:

$$
\begin{aligned}
U &= \left(\int_{\nu\in\mathcal{V}} \left(\varphi\left(\nu\right)\right)^{\frac{\sigma-1}{\sigma}} \left(YP^{\sigma-1}p(\nu)^{-\sigma}\varphi(\nu)^{\sigma-1}\right)^{\frac{\sigma-1}{\sigma}} d\nu\right)^{\frac{\sigma}{\sigma-1}(1-\rho)} L(\mathcal{V})^{\rho} \\
&= \left(YP^{\sigma-1} \underbrace{\left(\int_{\nu\in\mathcal{V}} \left(\frac{p(\nu)}{\varphi(\nu)}\right)^{1-\sigma} d\nu\right)^{\frac{\sigma}{\sigma-1}}}_{P^{-\sigma}}\right)^{1-\rho} L(\mathcal{V})^{\rho} \\
&= \left(Y\cdot P^{\sigma-1}\cdot P^{-\sigma}\right)^{1-\rho} L\left(\nu\right)^{\rho} = \left(\frac{Y}{P}\right)^{1-\rho} L(\mathcal{V})^{\rho} \\
&= \left((1-\rho)w\left(T-\overbrace{\int_0^D \mu(\nu)d\nu}^{\tau(\nu)}\right)\overbrace{\left(\int_0^D \left(\frac{p(\nu)}{\varphi(\nu)}\right)^{1-\sigma} d\nu\right)^{\frac{1}{\sigma-1}}}^{P^{-1}}\right)^{1-\rho} \left(\rho\left(T-\int_0^D \mu(\nu)d\nu\right)\right)^{\rho}
\end{aligned}
$$

There exists an optimal cutoff variety ν_D. The FOC for variety is then:

$$
\begin{aligned}
\frac{\partial U}{\partial D} &= L(D)^{\rho}(1-\rho)\left(\frac{Y}{P}\right)^{-\rho}\left(P^{-1}\frac{\partial Y}{\partial D}+Y\frac{\partial P^{-1}}{\partial D}\right)+\left(\frac{Y}{P}\right)^{1-\rho}\rho L(D)^{\rho-1}\frac{\partial L(D)}{\partial D} \\
&= L(D)^{\rho-1}\left(\frac{Y}{P}\right)^{1-\rho}\rho\left(\frac{1}{w}\frac{\partial Y}{\partial D}+\frac{1}{w}\frac{\partial P^{-1}}{\partial D}PY+\frac{\partial L(D)}{\partial D}\right) \\
&= \lambda\left[-(1-\rho)w\mu(\nu_D)+\frac{YP^{\sigma-1}}{\sigma-1}\left(\frac{p(\nu_D)}{\varphi(\nu_D)}\right)^{1-\sigma}-\rho w\mu(\nu_D)\right] \\
&= \lambda\left[-w\mu(\nu_D)+\frac{A(D)\left(\frac{p(\nu_D)}{\varphi(\nu_D)}\right)^{1-\sigma}}{\sigma-1}\right]
\end{aligned}
$$

Substituting these expressions into the utility function, we can follow Bronnenberg (2015) and derive that the optimal cutoff variety ν_D in the set of varieties D satisfies the following condition:

$$
\begin{aligned}
\frac{A(D)\left(\frac{p(\nu_D)}{\varphi(\nu_D)}\right)^{1-\sigma^F}}{\sigma^F-1} &= w\mu(\nu_D) \qquad (32)\\
A(D) = YP^{\sigma^F-1} &= (1-\rho)w\left(T-\tau(D)\right)\left(\int_{\nu\in D}\left(\frac{p^F(\nu)}{\varphi^F(\nu)}\right)^{1-\sigma^F} d\nu\right)^{-1}
\end{aligned}
$$

The FOC for variety can be derived similarly when allowing for leisure technology θ as follows:

$$
\begin{aligned}
X(D) &= \frac{1-\rho}{1-\rho+\theta\rho}\frac{w(T-\tau(D))+K}{P} \\
U &= \left(\frac{Y}{P}\right)^{1-\rho} L(D)^{\theta\rho} \\
\frac{\partial U}{\partial D} &= L(D)^{\theta\rho}(1-\rho)\left(\frac{Y}{P}\right)^{-\rho}\left(P^{-1}\frac{\partial Y}{\partial D}+Y\frac{\partial P^{-1}}{\partial D}\right)+\left(\frac{Y}{P}\right)^{1-\rho}\theta\rho L(D)^{\theta\rho-1}\frac{\partial L(D)}{\partial D} \\
&= L(D)^{\theta\rho-1}\left(\frac{Y}{P}\right)^{1-\rho}\theta\rho\left(\frac{1}{w}\frac{\partial Y}{\partial D}+\frac{1}{w}\frac{\partial P^{-1}}{\partial D}PY+\frac{\partial L(D)}{\partial D}\right) \\
&= \lambda\left[-w\mu(\nu_D)+\frac{A(D)\left(\frac{p(\nu_D)}{\varphi(\nu_D)}\right)^{1-\sigma}}{\sigma-1}\right] \\
Y &= w(T-\tau(D)-L)+K = \frac{1-\rho}{1-\rho+\theta\rho}\left(w(T-\tau(D))+K\right), A(D)=YP^{\sigma-1}
\end{aligned}
$$

Next, we assume that firms have identical taste parameters φ^F equal to one. This is simply done for expositional purposes and allows us to index each firm by their price index rather their taste-adjusted price index $\frac{P_f^F}{\varphi_f^F}$. We also assume that P_f^F lies on a continuum $[\underline{P^F}, \overline{P^F}]$.[38] We assume that the shopping cost at each firm is a function of the price index P_f^F in each firm, with $\mu'(P_f^F) < 0$. This is consistent with the fact that firms with larger variety and lower prices, i.e. lower P_f^F, tend to be located farther away from consumers relative to firms with lower variety and higher prices, due to higher costs of land in areas with higher population density and the need for larger square footage stores in order to stock higher variety and lower prices through economies of scale. We can show that if $\mu'(P_f^F)$ is small enough in absolute value, there exists a unique cutoff firm P^{F*} that satisfies the following condition using equation (17), such that consumers only buy from firms that have a price index within the set $[\underline{P^F}, P^{F*}]$:

$$
\left(P^{F*}\right)^{1-\sigma^F}\frac{1}{\int_{\underline{P^F}}^{P^{F*}}\left(P_f^F\right)^{1-\sigma^F}dP_f^F}\frac{(1-\rho)\left(w\left(T-\int_{\underline{P^F}}^{P^{F*}}\mu(p)dp\right)+K\right)}{w(\sigma^F-1)}-\mu(P^{F*})=0 \quad (33)
$$

[38] Assuming the firms lie on a continuum indexed by P_f^F allows for higher analytical tractability and the use of integrals. We can also allow for discrete P_f^F and use summations instead with similar intuitions for all of our derivations.

Uniqueness + existence of P^{F*}

$$f(P^F) = \underbrace{\left(P^F\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{P^F} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^F} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} - \mu(P^F)$$

$$\text{Assume } f(\underline{P^F}) = \left(\underline{P^F}\right)^{1-\sigma^F} \frac{1}{\left(\underline{P^F}\right)^{1-\sigma^F}} \frac{(1-\rho)\left(w\left(T - \mu(\underline{P^F})\right) + K\right)}{w(\sigma^F - 1)} - \mu(\underline{P^F}) > 0$$

$$\text{Assume } f(\overline{P^F}) = \left(\overline{P^F}\right)^{1-\sigma^F} \frac{1}{\int_{\underline{P^F}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} \frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{\overline{P^F}} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)} - \mu(\overline{P^F}) < 0$$

$$\begin{aligned} f'(P^F) &= (B^{-1}C)(1-\sigma^F)\left(P^F\right)^{-\sigma^F} \qquad < 0 \\ &+ (AC)\left(-B^{-2}\right)\left(P^F\right)^{1-\sigma^F} \qquad < 0 \\ &+ (AB^{-1})\frac{1-\rho}{\sigma^F - 1}\left(-\mu(P^F)\right) \qquad < 0 \\ &- \mu'(P^F) \qquad > 0 \\ &< 0 \qquad \text{if } \mu'(P^F) < 0 \text{ and is not too large in absolute value} \end{aligned}$$

E.2 Household Retail Concentration

Given our derivations, the market share for household i for each firm can be written as follows:

$$S_{fi}^F = \frac{P_f^F x^F(\nu)}{Y} = \begin{cases} \left(P\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F - 1} & \text{if } \frac{YP^{\sigma^F-1}\left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}}{w(\sigma^F - 1)} \geq \mu(P_f^F) \\ 0 & \text{otherwise} \end{cases}$$

Given our simplifying assumptions, we can further simplify the market share expression and write the household retail HHI for household i as follows:

$$H_i = \int_f \left(S_{fi}^F\right)^2 df = \frac{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{2\left(1-\sigma^F\right)} dP_f^F}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \tag{34}$$

We can then show some comparative statics of household retail HHI in response to changes in various parameters. First, we show that any parameters, in this case denoted by t, that decrease the cutoff firm price index P^{F*} will increase household retail HHI. This is intuitive since the set of retailers that households consume from, $[\underline{P^F}, P^{F*}]$, will decrease as

P^{F*} decreases.

$$\frac{dP^{F*}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0 \tag{35}$$

$$\begin{aligned}
\frac{dH_i}{dt} &= \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \\
&\cdot \left(-2\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{-3} \left(\left(P^{F*}\right)^{1-\sigma^F} \frac{dP^{F*}}{dt} + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial \left(P_f^F\right)^{1-\sigma^F}}{\partial t} dP_f^F\right)\right) \\
&+ \frac{1}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \left(\left(\left(P^{F*}\right)^{1-\sigma^F}\right)^2 \frac{dP^{F*}}{dt} + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2}{\partial t} dP_f^F\right) \\
&= \frac{\left(P^{F*}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \left(\frac{2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} - \left(P^{F*}\right)^{1-\sigma^F}\right) \left(-\frac{dP^{F*}}{dt}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{dP^{F*}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0 \text{ if } &\frac{2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} > \left(P^{F*}\right)^{1-\sigma^F} \\
\text{Jensen's inequality} \Rightarrow &\int \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \geqslant \left(\int \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2 \\
\text{WTS } 2 &\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F > \left(P^{F*}\right)^{1-\sigma^F} \int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F
\end{aligned}$$

$$\begin{aligned}
&2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \geqslant 2\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2 \\
&> \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2 \geq \left(P^{F*}\right)^{1-\sigma^F} \int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \\
&\because \int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \geq \left(P^{F*}\right)^{1-\sigma^F}
\end{aligned}$$

Given this fact, we can use equation (33) to derive comparative statics of the cutoff firm price index P^{F*} in response to changes in various parameters to see how these parameters change household retail HHI. These comparative statics can then be compared against our empirical findings.

First, consider changes on the demand side. Let the shopping cost be $\mu(P^F) = t\delta(P^F)$, where t is the time cost per distance traveled and $\delta(P^F)$ is the distance of the consumer from each firm indexed by P^F.

Extensive margin (P^{F*}, Comparative Statics)

$$f(P^{F*}) = \underbrace{\left(P^{F*}\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^{F*}} t\delta(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} - t\delta(P^{F*}) = 0$$

$$\begin{aligned}
\frac{dP^{F*}}{dt} &= -\frac{\frac{\partial f(P^{F*},t)}{\partial t}}{\frac{\partial f(P^{F*},t)}{\partial P^F}} \\
&= \frac{\delta(P^{F*}) + (AB^{-1})\frac{1-\rho}{\sigma^F-1}\left(\int_{\underline{P^F}}^{P^{F*}} \delta(p)dp\right)}{\frac{\partial f(P^{F*},t)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},t)}{\partial t} > 0, \frac{\partial f(P^{F*},t)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{dt} < 0
\end{aligned}$$

$$\begin{aligned}
\frac{dP^{F*}}{dw} &= -\frac{\frac{\partial f(P^{F*},w)}{\partial w}}{\frac{\partial f(P^{F*},w)}{\partial P^F}} \\
&= \frac{(AB^{-1})\frac{1-\rho}{\sigma^F-1}Kw^{-2}}{\frac{\partial f(P^{F*},w)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},w)}{\partial w} > 0, \frac{\partial f(P^{F*},\rho)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{dw} < 0
\end{aligned}$$

$$\begin{aligned}
\frac{dP^{F*}}{d\rho} &= -\frac{\frac{\partial f(P^{F*},\rho)}{\partial \rho}}{\frac{\partial f(P^{F*},\rho)}{\partial P^F}} \\
&= \frac{(AB^{-1})\frac{1}{w(\sigma^F-1)}\left(w\left(T - \int_{\underline{P^F}}^{P^{F*}} t\delta(p)dp\right) + K\right)}{\frac{\partial f(P^{F*},\rho)}{\partial P^F}} \\
&- \frac{\partial f(P^{F*},\rho)}{\partial \rho} > 0, \frac{\partial f(P^{F*},\rho)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\rho} < 0
\end{aligned}$$

$$f(P^{F*}) = \underbrace{\left(P^{F*}\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{P^{F*}} t\delta(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} \underbrace{\frac{1}{1-\rho+\theta\rho}}_{D^{-1}} - t\delta(P^{F*})$$

$$= 0$$

$$\frac{dP^{F*}}{d\theta} = -\frac{\frac{\partial f(P^{F*},\theta)}{\partial \theta}}{\frac{\partial f(P^{F*},\theta)}{\partial P^F}}$$

$$= \frac{\left(AB^{-1}C\right)\left(D^{-2}\right)\rho}{\frac{\partial f(P^{F*},t)}{\partial P^F}}$$

$$-\frac{\partial f(P^{F*},\theta)}{\partial \theta} > 0, \frac{\partial f(P^{F*},\theta)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\theta} < 0$$

$$\text{Let } \mu = \mu(N,\beta) \text{ and } \frac{\partial \mu}{\partial \beta} < 0$$

$$\frac{dP^{F*}}{d\beta} = -\frac{\frac{\partial f(P^{F*},\beta)}{\partial \beta}}{\frac{\partial f(P^{F*},\beta)}{\partial P^F}}$$

$$= \frac{\frac{\partial \mu}{\partial \beta} + AB^{-1}\frac{1-\rho}{\sigma^F-1}\int_{\underline{P^F}}^{P^{F*}} \frac{\partial \mu}{\partial \beta} dn}{\frac{\partial f(P^{F*},\beta)}{\partial P^F}}$$

$$-\frac{\partial f(P^{F*},\beta)}{\partial \beta} < 0, \frac{\partial f(P^{F*},\beta)}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\beta} > 0$$

$$\frac{dP^{F*}}{d\underline{P^F}} = -\frac{\frac{\partial f(P^{F*},\underline{P^F})}{\partial \underline{P^F}}}{\frac{\partial f(P^{F*},\underline{P^F})}{\partial P^F}}$$

$$= \frac{-AB^{-1}\frac{1-\rho}{\sigma^F-1}\mu(\underline{P^F}) - AC^{-1}(B^{-2})\underline{P^F}^{1-\sigma^F}}{\frac{\partial f(P^{F*},\beta)}{\partial P^F}}$$

$$-\frac{\partial f(P^{F*},\underline{P^F})}{\partial \underline{P^F}} < 0, \frac{\partial f(P^{F*},\underline{P^F})}{\partial P^F} < 0 \Rightarrow \frac{dP^{F*}}{d\underline{P^F}} > 0$$

$$
\begin{aligned}
H_i &= \int_f \left(S_{fi}^F\right)^2 df = \frac{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{2\left(1-\sigma^F\right)} dP_f^F}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \\
\frac{dH_i}{d\underline{P^F}} &= \int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \\
&\quad \cdot \left(-2\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{-3} \left(\left(P^{F*}\right)^{1-\sigma^F} \frac{dP^{F*}}{d\underline{P^F}} - \left(\underline{P^F}\right)^{1-\sigma^F} + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial \left(P_f^F\right)^{1-\sigma^F}}{\partial \underline{P^F}} dP_f^F\right)\right) \\
&\quad + \frac{1}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \left(\left(\left(P^{F*}\right)^{1-\sigma^F}\right)^2 \frac{dP^{F*}}{d\underline{P^F}} - \left(\left(\underline{P^F}\right)^{1-\sigma^F}\right)^2 + \int_{\underline{P^F}}^{P^{F*}} \frac{\partial \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2}{\partial \underline{P^F}} dP_f^F\right) \\
&= -\frac{\left(P^{F*}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \underbrace{\left(\frac{2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} - \left(P^{F*}\right)^{1-\sigma^F}\right)}_{A_1} \frac{dP^{F*}}{d\underline{P^F}} \\
&\quad + \frac{\left(\underline{P^F}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \underbrace{\left(\frac{2\int_{\underline{P^F}}^{P^{F*}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^F}}^{P^{F*}} \left(\underline{P^F}\right)^{1-\sigma^F} dP_f^F} - \left(\underline{P^F}\right)^{1-\sigma^F}\right)}_{A_2}
\end{aligned}
$$

$$
A_1 \geq A_2 > 0,\ \frac{dP^{F*}}{d\underline{P^F}} > \left(\frac{P^{F*}}{\underline{P^F}}\right)^{\sigma^F-1} \frac{A_2}{A_1} \Rightarrow \frac{dH_i}{d\underline{P^F}} < 0
$$

Let $P^F = P^F(\alpha)$ and $f(P^F, \alpha) = \left(P^F(\alpha)\right)^{1-\sigma^F}$, $\dfrac{dP^F}{d\alpha} < 0 \Rightarrow \dfrac{df}{d\alpha} > 0$

$$
\begin{aligned}
H_i &= \frac{\int (f(p,\alpha))^2 dp}{\left(\int f(p,\alpha) dp\right)^2} \\
\frac{dH_i}{d\alpha} &= \frac{1}{\left(\int f dp\right)^2} \int 2f \frac{df}{d\alpha} + \int f^2 \left(-2\left(\int f\right)^{-3} \int \frac{df}{d\alpha}\right) \\
&= \frac{2}{\left(\int f dp\right)^2} \left(\int f \frac{df}{d\alpha} - \int f^2 \left(\int f\right)^{-1} \int \frac{df}{d\alpha}\right) > 0 \text{ if} \\
& \int f \int f \frac{df}{d\alpha} - \int f^2 \int \frac{df}{d\alpha} > 0 \\
& \int f \frac{df}{d\alpha} > \frac{\int f^2 \int \frac{df}{d\alpha}}{\int f} \geq \frac{(\int f)^2 \int \frac{df}{d\alpha}}{\int f} = \int f \int \frac{df}{d\alpha}
\end{aligned}
$$

where we use Jensen's inequality for the $\geq$ condition

E.3 Welfare

$$
\begin{aligned}
U(X(\mathcal{V}), L(\mathcal{V})) &= X(\mathcal{V})^{1-\rho} L(\mathcal{V})^{\rho} \\
X(\mathcal{V}) &= \left(\int_{\nu \epsilon \mathcal{V}} (\varphi(\nu) x(\nu))^{\frac{\sigma-1}{\sigma}} \, dv \right)^{\frac{\sigma}{\sigma-1}} \\
\tau(D) &= t \int_0^D d(\nu) d\nu \\
\mu(\nu) &= t d(\nu) \\
U &= \left(\frac{Y}{P} \right)^{1-\rho} L\,(D)^{\rho} \\
\frac{Y}{P} &= (1-\rho) \left(w \left(T - \overbrace{\int_0^D \mu(\nu) d\nu}^{\tau(D) \equiv} \right) + K \right) \overbrace{\left(\int_0^D \left(\frac{p(\nu)}{\varphi(\nu)} \right)^{1-\sigma^F} d\nu \right)^{\frac{1}{\sigma^F - 1}}}^{P^{-1}} \\
L(D) &= \frac{\rho \left(w \left(T - \int_0^D \mu(\nu) d\nu \right) + K \right)}{w}
\end{aligned}
$$

Applying the envelope theorem, we have

$$
\begin{aligned}
\left. \frac{dU}{dt} \right|_{D=D^*} &= (1-\rho) X(D)^{-\rho} \left[-(1-\rho) w \int_0^D d(\nu) d\nu P^{-1} \right] L(D)^{\rho} \\
&\quad + X(D)^{1-\rho} \rho L(D)^{\rho-1} \left[-\rho \int_0^D d(\nu) d\nu \right] < 0 \\
&= -X(D)^{-\rho} L(D)^{\rho-1} \int_0^D d(\nu) d\nu \left[(1-\rho)^2 w P^{-1} \underbrace{L(D)}_{\rho(T-\tau(D))} + \rho^2 \underbrace{X(D)}_{(1-\rho)w(T-\tau(D))P^{-1}} \right] \\
&= -X(D)^{-\rho} L(D)^{\rho-1} \rho \int_0^D d(\nu) d\nu < 0
\end{aligned}
$$

$$
\begin{aligned}
U &= P^{\rho-1} (1-\rho)^{1-\rho} \rho^{\rho} \left[w^{1-\rho} (T - \tau(D)) + K w^{-\rho} \right] \\
\left. \frac{dU}{dw} \right|_{D=D^*} &= P^{\rho-1} (1-\rho)^{1-\rho} \rho^{\rho} \left[(1-\rho) w^{-\rho} (T - \tau(D)) - K \rho w^{-\rho-1} \right] \\
&= P^{\rho-1} (1-\rho)^{1-\rho} \rho^{\rho} w^{-\rho-1} \left[(1-\rho) w (T - \tau(D) + K - K) - \rho K \right] \\
&= P^{\rho-1} (1-\rho)^{1-\rho} \rho^{\rho} w^{-\rho-1} (Y - K) > 0 \\
&= P^{\rho-1} (1-\rho)^{1-\rho} \rho^{\rho} w^{-\rho} (T - \tau(D) - L(D)) > 0
\end{aligned}
$$

$$
\begin{aligned}
lnU &= (1-\rho)ln\left(\frac{Y}{P}\right) + \rho lnL(D) \\
\left.\frac{dlnU}{d\rho}\right|_{D=D^*} &= -ln\left(\frac{Y}{P}\right) + (1-\rho)\frac{1}{\frac{Y}{P}}\left(-w\left(T-\tau(D)\right)^{\rho-1}\right) + lnL(D) + \rho\frac{1}{L(D)}(T-\tau(D)) \\
&= -ln(\frac{Y}{P}) - 1 + lnL(D) + 1 = -ln\left(\frac{Y}{P}\right) + lnL(D) > 0 \text{ if } L(D) > \frac{Y}{P}
\end{aligned}
$$

$$
\begin{aligned}
\text{If } U\left(X(\mathcal{V}), L(\mathcal{V})\right) &= X(\mathcal{V})^{1-\rho}L(\mathcal{V})^{\theta\rho} \\
lnU &= (1-\rho)ln\left(\frac{Y}{P}\right) + \theta\rho lnL(D) \\
\left.\frac{dlnU}{d\theta}\right|_{D=D^*} &= \rho lnL(D) > 0
\end{aligned}
$$

$$
\begin{aligned}
\text{Let } \mu = \mu(P^F, \beta) \text{ and } \frac{d\mu}{d\beta} < 0 \\
\left.\frac{dU}{d\beta}\right|_{D=D^*} = -(1-\rho)X(D)^{-\rho}(1-\rho)wP^{-1}L(D)^{\rho}\frac{d\mu}{d\beta} - X(D)^{1-\rho}\rho L(D)^{\rho-1}\rho\frac{d\mu}{d\beta} > 0
\end{aligned}
$$

Using the FOC for variety and the conditions required for the existence of a unique cutoff price index, we have

$$
\left.\frac{dU}{d\underline{P^F}}\right|_{D=D^*} = -\rho^{\rho}(1-\rho)^{1-\rho}\left(\frac{w}{P}\right)^{1-\rho}\left[-\mu(\underline{P^F}) + \frac{A(D)\left(\underline{P^F}\right)^{1-\sigma}}{w(\sigma-1)}\right] < 0
$$

Let $P^F = P^F(\alpha)$ and $f(P^F, \alpha) = P^F(\alpha)^{1-\sigma^F}$, $\dfrac{dP^F}{d\alpha} < 0 \Rightarrow \dfrac{df}{d\alpha} > 0$

$$\left.\frac{dU}{d\alpha}\right|_{D=D^*} = \left(\overbrace{(1-\rho)\left(w(T-\tau(D))+K\right)}^{Y(D)}\right)^{1-\rho} \left(\frac{\overbrace{\rho\left(w\left(T-\tau(D)\right)+K\right)}^{L(D)}}{w}\right)^{\rho} \frac{dP^{\rho-1}}{d\alpha} = ...$$

$$\frac{dP^{\rho-1}}{d\alpha} = (\rho-1)P^{\rho-2}\frac{dP}{d\alpha} = \underbrace{(\rho-1)}_{<0} P^{\rho-2} \underbrace{\frac{1}{1-\sigma^F}}_{<0} \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F}\right)^{\frac{\sigma^F}{1-\sigma^F}} \underbrace{\frac{d\int_{\underline{P^F}}^{P^{F*}} f(P^F,\alpha)dP^F}{d\alpha}}_{>0} > 0$$

$$P = \left(\int_0^D \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{\frac{1}{1-\sigma^F}}$$

E.4 Heterogeneity

If there is a group of consumers with $f'(P^F) > 0$, $f(\underline{P^F}) < 0$ and $f(\overline{P^F}) > 0$, i.e. $\mu'(P^F)$ large enough in absolute value $\Rightarrow \exists \underline{P^{F*}} s.t.\ f \in \left[\underline{P^{F*}}, \overline{P^F}\right]$

One group of consumers with $\overline{P^{F*}} \downarrow$ if $\frac{d\overline{P^{F*}}}{dt} < 0 \Rightarrow \frac{dH_i}{dt} > 0$, another group of consumers with $\underline{P^{F*}}^*$ $\uparrow$ if $\frac{d\underline{P^{F*}}}{dt} > 0 \Rightarrow \frac{dH_i}{dt} > 0$, then cross-variance $\uparrow$ and each group polarizes into their respective set of firms.

$\underline{\text{Uniqueness + existence of } \underline{P^{F*}}}$

$$f(P^F) = \underbrace{\left(P^F\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{P^F}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{P^F}^{\overline{P^F}} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} - \mu(P^F)$$

$$\text{Assume } f(\underline{P^F}) = \left(\underline{P^F}\right)^{1-\sigma^F} \frac{1}{\int_{\underline{P^F}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} \frac{(1-\rho)\left(w\left(T - \int_{\underline{P^F}}^{\overline{P^F}} \mu(p)dp\right) + K\right)}{w(\sigma^F - 1)} - \mu(\underline{P^F}) < 0$$

$$\text{Assume } f(\overline{P^F}) = \left(\overline{P^F}\right)^{1-\sigma^F} \frac{1}{\left(\overline{P^F}\right)^{1-\sigma^F}} \frac{(1-\rho)\left(w\left(T - \mu(\overline{P^F})\right) + K\right)}{w(\sigma^F - 1)} - \mu(\overline{P^F}) > 0$$

$$\begin{aligned} f'(P^F) &= (B^{-1}C)(1-\sigma^F)\left(P^F\right)^{-\sigma^F} \qquad < 0 \\ &+ (AC)\left(B^{-2}\right)\left(P^F\right)^{1-\sigma^F} \qquad > 0 \\ &+ (AB^{-1})\frac{1-\rho}{\sigma^F - 1}\left(\mu(P^F)\right) \qquad > 0 \\ &- \mu'(P^F) \qquad > 0 \\ &> 0 \qquad \text{if } \mu'(P^F) < 0 \text{ and is large enough in absolute value} \end{aligned}$$

$$H_i = \frac{\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{2\left(1-\sigma^F\right)} dP_f^F}{\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \qquad f \in \left[\underline{P^{F*}}, \overline{P^F}\right]$$

$$\begin{aligned} \frac{dH_i}{dt} &= \int_{\underline{P^{F*}}}^{\overline{P^F}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \\ &\cdot \left(-2\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{-3} \left(-\left(\underline{P^{F*}}\right)^{1-\sigma^F} \frac{d\underline{P^{F*}}}{dt} + \int_{\underline{P^{F*}}}^{\overline{P^F}} \frac{\partial \left(P_f^F\right)^{1-\sigma^F}}{\partial t} dP_f^F\right)\right) \\ &+ \frac{1}{\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \left(-\left(\left(\underline{P^{F*}}\right)^{1-\sigma^F}\right)^2 \frac{d\underline{P^{F*}}}{dt} + \int_{\underline{P^{F*}}}^{\overline{P^F}} \frac{\partial \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2}{\partial t} dP_f^F\right) \\ &= \frac{\left(\underline{P^{F*}}\right)^{1-\sigma^F}}{\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2} \left(\frac{2\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} - \left(\underline{P^{F*}}\right)^{1-\sigma^F}\right) \frac{d\underline{P^{F*}}}{dt} \end{aligned}$$

$$\frac{d\underline{P^{F*}}}{dt} > 0 \Rightarrow \frac{dH_i}{dt} > 0 \text{ if } \frac{2\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F}{\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F} > \left(\underline{P^{F*}}\right)^{1-\sigma^F}$$

$$\text{Jensen's inequality} \Rightarrow \int \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \geqslant \left(\int \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2$$

$$\text{WTS } 2\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F > \left(\underline{P^{F*}}\right)^{1-\sigma^F} \int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F$$

$$2\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(\left(P_f^F\right)^{1-\sigma^F}\right)^2 dP_f^F \geqslant 2\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2$$

$$> \left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^2 \geq \left(\underline{P^{F*}}\right)^{1-\sigma^F} \int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F$$

$$\because \int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F \geq \left(\underline{P^{F*}}\right)^{1-\sigma^F}$$

Extensive margin ($\underline{P^{F*}}$, Comparative Statics)

$$f(\underline{P^{F*}}) = \underbrace{\left(\underline{P^{F*}}\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T - \int_{\underline{P^{F*}}}^{\overline{P^F}} t\delta(p)dp\right) + K\right)}{w(\sigma^F - 1)}}_{C} - t\delta(\underline{P^{F*}}) = 0$$

$$\begin{aligned}
\frac{d\underline{P^{F*}}}{dt} &= -\frac{\frac{\partial f(\underline{P^{F*}},t)}{\partial t}}{\frac{\partial f(\underline{P^{F*}},t)}{\partial P^F}} \\
&= \frac{\delta(\underline{P^{F*}}) + (AB^{-1})\,\frac{1-\rho}{\sigma^F-1}\left(\int_{\underline{P^{F*}}}^{\overline{P^F}} \delta(p)dp\right)}{\frac{\partial f(\underline{P^{F*}},t)}{\partial P^F}} \\
&- \frac{\partial f(\underline{P^{F*}},t)}{\partial t} > 0, \frac{\partial f(\underline{P^{F*}},t)}{\partial P^F} > 0 \Rightarrow \frac{d\underline{P^{F*}}}{dt} > 0
\end{aligned}$$

$$\begin{aligned}
\frac{d\underline{P^{F*}}}{dw} &= -\frac{\frac{\partial f(\underline{P^{F*}},w)}{\partial w}}{\frac{\partial f(\underline{P^{F*}},w)}{\partial P^F}} \\
&= \frac{(AB^{-1})\,\frac{1-\rho}{\sigma^F-1} K w^{-2}}{\frac{\partial f(\underline{P^{F*}},w)}{\partial P^F}} \\
&- \frac{\partial f(\underline{P^{F*}},w)}{\partial w} > 0, \frac{\partial f(\underline{P^{F*}},\rho)}{\partial P^F} > 0 \Rightarrow \frac{d\underline{P^{F*}}}{dw} > 0
\end{aligned}$$

$$\begin{aligned}
\frac{d\underline{P^{F*}}}{d\rho} &= -\frac{\frac{\partial f(\underline{P^{F*}},\rho)}{\partial \rho}}{\frac{\partial f(\underline{P^{F*}},\rho)}{\partial P^F}} \\
&= \frac{(AB^{-1})\,\frac{1}{w(\sigma^F-1)}\left(w\left(T - \int_{\underline{P^{F*}}}^{\overline{P^F}} t\delta(p)dp\right) + K\right)}{\frac{\partial f(\underline{P^{F*}},\rho)}{\partial P^F}} \\
&- \frac{\partial f(\underline{P^{F*}},\rho)}{\partial \rho} > 0, \frac{\partial f(\underline{P^{F*}},\rho)}{\partial P^F} > 0 \Rightarrow \frac{d\underline{P^{F*}}}{d\rho} > 0
\end{aligned}$$

$$f(\underline{P^{F*}}) = \underbrace{\left(\underline{P^{F*}}\right)^{1-\sigma^F}}_{A} \underbrace{\frac{1}{\int_{\underline{P^{F*}}}^{\overline{P^F}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F}}_{B^{-1}} \underbrace{\frac{(1-\rho)\left(w\left(T-\int_{\underline{P^{F*}}}^{\overline{P^F}} t\delta(p)dp\right)+K\right)}{w(\sigma^F-1)}}_{C} \underbrace{\frac{1}{1-\rho+\theta\rho}}_{D^{-1}} - t\delta(\underline{P^{F*}})$$

$$= 0$$

$$\frac{d\underline{P^{F*}}}{d\theta} = -\frac{\frac{\partial f(\underline{P^{F*}},\theta)}{\partial\theta}}{\frac{\partial f(\underline{P^{F*}},\theta)}{\partial P^F}}$$

$$= \frac{\left(AB^{-1}C\right)\left(D^{-2}\right)\rho}{\frac{\partial f(\underline{P^{F*}},t)}{\partial P^F}}$$

$$- \frac{\partial f(\underline{P^{F*}},\theta)}{\partial\theta} > 0, \frac{\partial f(\underline{P^{F*}},\theta)}{\partial P^F} > 0 \Rightarrow \frac{d\underline{P^{F*}}}{d\theta} > 0$$

$$\text{Let } \mu = \mu(N,\beta) \text{ and } \frac{\partial\mu}{\partial\beta} < 0$$

$$\frac{d\underline{P^{F*}}}{d\beta} = -\frac{\frac{\partial f(\underline{P^{F*}},\beta)}{\partial\beta}}{\frac{\partial f(\underline{P^{F*}},\beta)}{\partial P^F}}$$

$$= \frac{\frac{\partial\mu}{\partial\beta} + AB^{-1}\frac{1-\rho}{\sigma^F-1}\int_{\underline{P^{F*}}}^{\overline{P^F}} \frac{\partial\mu}{\partial\beta}dn}{\frac{\partial f(\underline{P^{F*}},\beta)}{\partial P^F}}$$

$$- \frac{\partial f(\underline{P^{F*}},\beta)}{\partial\beta} < 0, \frac{\partial f(\underline{P^{F*}},\beta)}{\partial P^F} > 0 \Rightarrow \frac{d\underline{P^{F*}}}{d\beta} < 0$$

$$\underline{P^{F*}} > \underline{P^F} \Rightarrow \frac{d\underline{P^{F*}}}{d\underline{P^F}} = -\frac{\frac{\partial f(\underline{P^{F*}},\underline{P^F})}{\partial\underline{P^F}}}{\frac{\partial f(\underline{P^{F*}},\underline{P^F})}{\partial P^F}} = 0$$

$$\text{Let } P^F = P^F(\alpha) \text{ and } f(P^F,\alpha) = P^F(\alpha)^{1-\sigma^F},\ \frac{dP^F}{d\alpha} < 0 \Rightarrow \frac{df}{d\alpha} > 0$$

$$H_i = \frac{\int (f(p,\alpha))^2 dp}{\left(\int f(p,\alpha)dp\right)^2}$$

$$\frac{dH_i}{d\alpha} = \frac{1}{\left(\int f dp\right)^2}\int 2f\frac{df}{d\alpha} + \int f^2\left(-2\left(\int f\right)^{-3}\int\frac{df}{d\alpha}\right)$$

$$= \frac{2}{\left(\int f dp\right)^2}\left(\int f\frac{df}{d\alpha} - \int f^2\left(\int f\right)^{-1}\int\frac{df}{d\alpha}\right) > 0 \text{ if}$$

$$\int f\int f\frac{df}{d\alpha} - \int f^2\int\frac{df}{d\alpha} > 0$$

$$\int f\frac{df}{d\alpha} > \frac{\int f^2\int\frac{df}{d\alpha}}{\int f} \geq \frac{(\int f)^2\int\frac{df}{d\alpha}}{\int f} = \int f\int\frac{df}{d\alpha}$$

where we use Jensen's inequality for the $\geq$ condition

E.5 Markups

What are the implications of our empirical results for market power? First, consider a general profit maximization problem for each firm f:

$$\max_{P_f^F}\Pi_f^F = S_f^F\int_i Y_i di - C_f$$

Firm f chooses its firm price index P_f^F given its cost function C_f and its revenue is driven by demand equations that determine its market share S_f^F. The FOC is then given by:

$$\frac{\partial\Pi_f^F}{\partial P_f^F} = \int_i Y_i di\frac{\partial S_f^F}{\partial P_f^F} - \frac{\partial C_f}{\partial P_f^F} = 0$$

Next, we have the following identity which relates the elasticity of the firm's residual demand elasticity of demand ε_f^F to the market share elasticity with respect to price:

$$\varepsilon_f^F = 1 - \frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F}$$

The FOC then gives us the following equation for the firm-level markup $\mathcal{M}_f^F$ for firm f, which is defined as the price index P_f^F divided by the marginal cost MC_f, as a function of

ε_f^F, which is the same across all products.[39]

$$
\begin{aligned}
\frac{\partial \Pi_f^F}{\partial P_f^F} &= \int_i Y_i di \frac{\partial S_f^F}{\partial P_f^F} - \frac{\partial C_f}{\partial P_f^F} = 0 \\
\frac{S_f^F \int_i Y_i di}{P_f^F}(1-\varepsilon_f^F) + \frac{\partial C_f}{\partial x_f}\frac{\partial x_f}{\partial P_f^F}\frac{P_f^F}{X_f^F}\frac{X_f^F}{P_f^F} &= 0 \\
X_f^F(1-\varepsilon_f^F) + MC_f \varepsilon_f^F \frac{X_f^F}{P_f^F} &= 0 \\
\mathcal{M}_f^F \equiv \frac{P_f^F}{mc_f} &= \frac{\varepsilon_f^F}{\varepsilon_f^F - 1}
\end{aligned}
$$

Note that X_f^F is the firm quantity index that equates revenue $S_f^F \int_i Y_i di$ to the product of the price index P_f^F and quantity index X_f^F.

How does a firm choose its price index? We show a more detailed supply-side setup following Hottman et al. (2016). For simplicity, consider the case where products are symmetric, such that firm f only chooses the number of products N_f and price p_f to determine its firm price index, i.e. we assume that products are symmetric such that products have identical taste parameters φ^U equal to one and identical prices p_f in each firm f. We can index each firm f by their price index P_f^F, which is a function of the number of products it sells N_f and price p_f, abstracting away from idiosyncratic tastes for each product and firm.

The firm price index for the cutoff firm P^{F*}, denoted also as the cutoff variety ν_D, and N_D products can be written more simply as follows:

$$
\begin{aligned}
P^{F*} &= p(\nu_D) = \left(\int_0^{N_D} p_D^{1-\sigma^U} dn\right)^{\frac{1}{1-\sigma^U}} = p_D N_D^{\frac{1}{1-\sigma^U}} \\
P_f^F &= p_f N_f^{\frac{1}{1-\sigma^U}} = \frac{p_f}{N_f^{\frac{1}{\sigma^U-1}}},\ P = \left(\int_{\underline{P^F}}^{P^{F*}} \left(P_f^F\right)^{1-\sigma^F} dP_f^F\right)^{\frac{1}{1-\sigma^F}}
\end{aligned} \tag{36}
$$

We can then formulate the profit-maximization problem as follows:

$$
\max_{N_f, p_f} \Pi_f^F = N_f p_f x_f - N_f A(x_f) - B(N_f) - H^F
$$

Let $A(x_f)$ be the variable cost function with respect to the quantities of each product x_f, and $B(N_f)$ be the variable cost function with respect to N_f, and H^F be the fixed cost. We then have the following FOC with respect to p_f:

[39]This is true even without imposing product symmetry and is a property of nested demand systems, as mentioned in Hottman et al. (2016). We show a more detailed supply-side setup that generates these equations in Appendix Section E.5.

$$\frac{\partial \Pi_f^F}{\partial p_f} = \int_i Y_i di \frac{\partial S_f^F}{\partial p_f} - \frac{\partial C_f}{\partial p_f} = 0$$
$$\frac{S_f^F \int_i Y_i di}{p_f}(1-\varepsilon_f^F) + N_f mc_f \partial x_f \frac{\partial x_f}{\partial p_f}\frac{p_f}{x_f}\frac{x_f}{p_f} = 0$$
$$N_f x_f (1-\varepsilon_f^F) + N_f mc_f \varepsilon_f^F \frac{X_f^F}{P_f^F} = 0$$
$$\mathcal{M}_f^F \equiv \frac{p_f}{mc_f} = \frac{\varepsilon_f^F}{\varepsilon_f^F - 1}$$

Equation 36 gives the price index when products are symmetric. Therefore, the following equation shows that the demand elasticity with respect to P_f^F is equal to the demand elasticity with respect to p_f:

$$N_f x_f p_f = X_f^F P_f^F$$
$$N_f^{\frac{-\sigma^U}{1-\sigma^U}} x_f = X_f^F$$
$$N_f^{\frac{-\sigma^U}{1-\sigma^U}} \frac{\partial x_f}{\partial p_f}\frac{p_f}{x_f} = \frac{\partial X_f^F}{\partial p_f}\frac{p_f}{x_f}$$
$$\epsilon_f \equiv \frac{\partial x_f}{\partial p_f}\frac{p_f}{x_f} = \frac{\partial X_f^F}{\partial P_f^F}\frac{P_f^F}{X_f^F} \equiv \varepsilon_f^F$$

We also have the following FOC with respect to N_f:

$$\frac{\partial \Pi_f^F}{\partial N_f} = \int_i Y_i di \frac{\partial S_f^F}{\partial N_f} - \frac{\partial C_f}{\partial N_f} = 0$$
$$\frac{S_f^F}{N_f}\int_i Y_i di \frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F}\frac{\partial P_f^F}{\partial N_f}\frac{N_f}{P_f^F} - \frac{\partial C_f}{\partial N_f} = 0$$
$$\frac{S_f^F}{N_f}\int_i Y_i di \frac{1-\varepsilon_f^F}{1-\sigma^U} - \frac{\partial C_f}{\partial N_f} = 0$$
$$N_f = \frac{S_f^F \int_i Y_i di \frac{\varepsilon_f^F - 1}{\sigma^U - 1}}{\frac{\partial C_f}{\partial N_f}}$$

We then show how the firm market share for the entire market S_f^F is related to the individual household market share s_{fi}^F. Recall that the market share of firm f for household

i S_{fi}^F can be written as follows:

$$S_{fi}^F = S_f^F(D_i) = \begin{cases} \left(P\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1} & \text{if } \frac{YP^{\sigma^F-1}\left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}}{w(\sigma^F-1)} \geq \mu_{fi}^F = \mu_i(P_f^F) \\ 0 & \text{otherwise} \end{cases}$$

$$P = P(D_i) = \left(\int_{f\in D_i}\left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F} dP_f^F\right)^{\frac{1}{1-\sigma^F}}$$

$$Y_i = Y(D_i) = (1-\rho)\left(w\left(T - \int_{f\in D_i}\mu_{fi}^F df\right) + K\right)$$

The market share, household price index, and income are all dependent on the specific set of firms D_i that household i consumes from, and D_i is a function of μ_{fi}^F by equation (17). We now allow for consumer heterogeneity more generally, allowing each household i to have a different shopping cost function μ_{fi}^F with probability density function $f(\mu_{fi}^F)$, where the shopping cost function is decreasing in P_f^F as mentioned previously. Since equation (17) is monotonic in μ_{fi}^F, for each firm f there exists a cutoff household $i*$ such that equation (17) holds. Household $i*$ then represents the marginal consumer for firm f. We can then write the market share S_f^F as follows:

$$\forall f,\ \exists i^* \text{ s.t. } \frac{Y(D_{i*})S_{fi}^F(D_{i*})}{\sigma^F-1} = w\mu_{fi*}^F$$

$$S_f^F = \frac{\int_i Y_i S_{fi}^F di}{\int_i Y_i di} = \frac{\int_0^{\mu_{fi*}^F} Y(D_i)S_f^F(D_i)f(\mu_{fi}^F)d\mu_{fi}^F}{\int_i Y_i di}$$

The firm market share elasticity can then be derived as follows:

$$\begin{aligned}\frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F} = &\frac{1}{\int_i Y_i di}\left[Y(D_{i*})S_f^F(D_{i*})f(\mu_{fi*}^F)\frac{1}{w(\sigma^F-1)}\frac{\partial Y(D_{i*})S_f^F(D_{i*})}{\partial P_f^F}\right.\\ &\left.+\int_0^{\mu_{fi*}^F}\frac{\partial Y(D_i)S_f^F(D_i)f(\mu_{fi}^F)}{\partial P_f^F}d\mu_{fi}^F\right]\frac{P_f^F}{S_f^F}\\ &-\int_i Y_i S_{fi}^F di\left[\left(\frac{1}{\int_i Y_i di}\right)^2\int_0^\infty\frac{\partial Y(D_i)f(\mu_{fi}^F)}{\partial P_f^F}d\mu_{fi}^F\right]\frac{P_f^F}{S_f^F}\end{aligned}$$

$$\frac{\partial Y(D_i)S_f^F(D_i)}{\partial P_f^F} = Y(D_i)\frac{\partial S_{fi}^F(D_i)}{\partial P_f^F} + S_f^F(D_i)\frac{\partial Y(D_i)}{\partial P_f^F}$$

For ease of exposition, assume that income Y_i is independent of P_f^F.[40] We can then

[40] As shown above, relaxing this assumption would allow the Y_i to shift as weights in response to changes in prices, which complicates the analysis without affecting the main intuition.

rewrite the market share elasticity more simply as follows:

$$\frac{\partial S_f^F}{\partial P_f^F}\frac{P_f^F}{S_f^F} = \frac{\partial S_f^F(D_{i*})}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_{i*})}\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F$$
$$+\int_0^{\mu_{fi*}^F}\frac{\partial S_f^F(D_i)}{\partial P_f^F}\frac{P_f^F}{S_f^F(D_i)}\frac{S_f^F(D_i)}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}d\mu_{fi}^F$$

Given the household market share S_{fi}^F, we can derive the following household market share elasticities under monopolistic competition and Bertrand competition:

$$S_{fi}^F = \left(P\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1}$$

$$\frac{\partial S_{fi}^F}{\partial P_f^F} = (P\varphi_f^F)^{\sigma^F-1}\frac{\partial\left(\frac{1}{P_f^F}\right)^{\sigma^F-1}}{\partial P_f^F} + \left(\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1}\frac{\partial P^{\sigma^F-1}}{\partial P_f^F}$$
$$= (P\varphi_f^F)^{\sigma^F-1}(1-\sigma^F)(P_f^F)^{-\sigma^F} + \left(\frac{\varphi_f^F}{P_f^F}\right)^{\sigma^F-1}\frac{\partial P^{\sigma^F-1}}{\partial P_f^F}$$

$$P = \left(\int_f \left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}\right)^{\frac{1}{1-\sigma^F}} \quad \text{if monopolistic competition}$$

$$P = \left(\sum_f \left(\frac{P_f^F}{\varphi_f^F}\right)^{1-\sigma^F}\right)^{\frac{1}{1-\sigma^F}} \quad \text{if Bertrand competition}$$

$$\frac{\partial P^{\sigma^F-1}}{\partial P_f^F} = \begin{cases} 0 & \text{if monopolistic competition} \\ (\sigma^F-1)P^{\sigma^F-2}\varphi^{\sigma^F-1}\left(\frac{P}{P_f^F}\right)^{\sigma^F} & \text{if Bertrand competition}\end{cases}$$

$$\frac{\partial S_{fi}^F}{\partial P_f^F}\frac{P_f^F}{S_{fi}^F} = \begin{cases} 1-\sigma^F & \text{if monopolistic competition} \\ (1-S_{fi}^F)(1-\sigma^F) & \text{if Bertrand competition}\end{cases}$$

Under monopolistic competition, the market share elasticity is only dependent on the elasticity of substitution σ^F. Under Bertrand competition, a change in the price index of firm f affects the aggregate price index P. The market share elasticity depends on both σ^F and the market share S_{fi}^F. A larger market share lowers the absolute value of the market share elasticity, lowering the elasticity of demand and raising the markup. Substituting the household market share elasticities into the market share elasticity and hence the demand elasticity, we can derive the markup under monopolistic competition and

Bertrand competition:

Monopolistic Competition

$$\mathcal{M}_f^F = \frac{\sigma^F + EM}{\sigma^F - 1 + EM} \tag{37}$$

$$EM = (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F = \frac{\left(S_f^F(D_{i*})\right)^2}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\frac{Y(D_{i*})}{w}$$

Bertrand Competition

$$\mathcal{M}_f^F = \frac{\sigma^F + EM - IM}{\sigma^F - 1 + EM - IM} \tag{38}$$

$$EM = (\sigma^F - 1)\frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\mu_{fi*}^F(1 - S_f^F(D_{i*}))$$

$$EM = \frac{S_f^F(D_{i*})}{S_f^F}\frac{Y(D_{i*})f(\mu_{fi*}^F)}{\int_i Y_i di}\frac{Y(D_{i*})}{w}S_f^F(D_{i*})(1 - S_f^F(D_{i*}))$$

$$IM = (\sigma^F - 1)\int_0^{\mu_{fi*}^F}\frac{1}{S_f^F}\frac{Y(D_i)f(\mu_{fi}^F)}{\int_i Y_i di}\left(S_f^F(D_i)\right)^2 d\mu_{fi}^F$$

Therefore, the markup differs from the standard CES markup $\frac{\sigma^F}{\sigma^F-1}$ in two ways. First, there is an additional extensive margin term EM that lowers the markup, since the existence of such marginal consumers increases the demand elasticity. This term is a function of several terms which include the share of firm sales from marginal consumers and the shopping cost at the margin. Note that there is also an additional term $1 - S_f^F(D_{i*})$ to account for the effect on the aggregate price index under Bertrand competition. This term is similar to the additional extensive margin term in Neiman and Vavra (2021). In contrast to their paper which focuses on taste heterogeneity across products, we allow for consumer heterogeneity to work through the shopping cost μ.

Second, there is an intensive margin term IM that increases markups due to Bertrand competition in addition to consumer heterogeneity. This term has also been highlighted in Feenstra et al. (2022), who again focus on taste heterogeneity across products. Intuitively, firms do not weight demand elasticities across consumers equally, but optimally use a greater weight on low-elasticity consumers, such that when consumer heterogeneity is present, they can charge higher markups and obtain higher profits, since the gains from charging higher markups to the lower elasticity consumers offsets the loss in demand on higher elasticity consumers. Crucially, this term increases as the variance of market shares across households increases, since firms can charge higher markups when there is a larger share of lower elasticity consumers.

The aggregate markup $\mathcal{M}$ is then the share-weighted average of all firm markups:

$$\mathcal{M} = \int_f S_f^F \mathcal{M}_f^F df$$

F Tables

Table F1: Change in cross-region variance, relative number of stores and markets, 2004-2019

	mean	sd	min	p25	p50	p75	max	count
Other chains								
Cross-region variance	5.26e-06	.0000535	-.000119	-1.23e-07	2.32e-08	3.65e-07	.00122	715
Number of stores	1.29	1.98	0	.5	.934	1.34	28	270
Number of markets	1.21	.97	.167	1	1	1	11	269
Top chains								
Cross-region variance	.0000849	.000604	-.0036	-5.82e-07	1.62e-06	.0000623	.00251	74
Number of stores	1.47	1.29	.157	.839	1.11	1.78	8.5	51
Number of markets	1.11	.636	.25	.727	1	1.22	4	51

Notes: This table summarizes the change in cross-region variance (in the number of stores), the relative number of stores and markets for each retail chain from 2004-2019, separating the distribution into two groups: the top 40 retail chains by revenue rank over the sample period, and other chains outside the top 40. The relative number of stores and markets is normalized to 1 in 2004q1.

Table F2: Change in cross-region variance, 2004-2019

	mean	sd	min	p25	p50	p75	max	count
Other chains	9.49e-06	.0000766	-.000728	-2.49e-08	1.58e-09	5.22e-07	.000948	664
Top chains	.000111	.000792	-.00416	-.0000224	.0000288	.000428	.00143	51

Notes: This table summarizes the change in cross-region variance for each retail chain from 2004-2019, separating the distribution into two groups: the top 40 retail chains by revenue rank over the sample period, and other chains outside the top 40.

Table F3: RZ-DOPD: Change in aggregate HHI, 2004-2015

	Category HHI	Category Variance	Household Variance	Region Variance
Survivor	0.05446	0.00363	-0.05547	0.02463
Within Survivor	0.02969	0.04744	-0.04799	0.03182
Between Survivor	0.02477	-0.04381	-0.00748	-0.00719
Entrants	0.00185	0.01687	-0.00378	-0.00218
Exiters	-0.02241	0.00302	0.01210	0.00089

Notes: This table decomposes the change in aggregate HHI following Appendix Section A.1. We also use a Dynamic Olley-Pakes decomposition (Melitz and Polanec 2015) on top of the RZ decomposition and apply it to each of the four terms that the aggregate HHI is composed of: Household-category HHI, household cross-category variance, regional cross-household variance, and national cross-region variance.

Table F4: Effect of Entry: Minimum Distance and Within Different Distance Thresholds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Household Retail Concentration							
	Supercenters				Clubs			
min_dis	-0.000399**				0.000703***			
	(0.000173)				(0.000158)			
num_5mi		0.00245**				-0.00520***		
		(0.00115)				(0.00158)		
num_10mi			0.000719				-0.00229**	
			(0.000581)				(0.000955)	
num_15mi				0.000515				-0.00141**
				(0.000395)				(0.000704)
Observations	1,616,203	1,616,203	1,616,203	1,616,203	467,181	467,181	467,181	467,181
R-squared	0.696	0.696	0.696	0.696	0.681	0.681	0.681	0.681
Prob > F	0.021	0.033	0.216	0.192	0.000	0.001	0.017	0.046
Household-Quarter FE	X	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X	X
Number of units	109224	109224	109224	109224	29187	29187	29187	29187
Number of clusters	109224	109224	109224	109224	29187	29187	29187	29187
BOTE	0.019	0.019	0.018	0.025	-0.086	-0.064	-0.057	-0.056

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Includes only HHs living in zip5 areas that have store entry within 20 miles. min_dis refers to the distance to the nearest supercenter or club, while num_5mi, num_10mi, and num_15mi refer to the number of stores within a 5-mile, 10-mile, and 15-mile radius from the zip-code centroid of each household. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table F5: Effects of Entry on Other Measures of Concentration

VARIABLES	(1) Num Retailers	(2) Within dept HHI	(3) Cross dept variance	(4) Cross HH variance	(5) Num Retailers	(6) Within dept HHI	(7) Cross dept variance	(8) Cross HH variance
NumSup	-0.229***	0.00281	-0.00542***	-0.000173				
	(0.0447)	(0.00244)	(0.00109)	(0.00263)				
NumClub3					0.246***	-0.0154***	-0.00134	-0.00264
					(0.0562)	(0.00302)	(0.00129)	(0.00325)
Observations	121,723	134,482	134,482	134,482	98,588	99,805	99,805	99,805
R-squared	0.695	0.667	0.506	0.627	0.703	0.667	0.510	0.632
Prob > F	0.000	0.249	0.000	0.947	0.000	0.000	0.299	0.416
Household-Quarter FE	X	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X	X
Number of units	9062	9371	9371	9371	6560	6323	6323	6323
Number of clusters	9062	9371	9371	9371	6560	6323	6323	6323
BOTE	0.167	0.112	0.297	-0.004	-0.180	-0.752	0.074	-0.050

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Includes only households living in zip5 with store entry during the period. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table F6: Effects of Entry on Number of Trips and Varieties per Trip

VARIABLES	(1) TripNum	(2) UPC per trip	(3) ProductGroup per trip	(4) Dept per trip	(5) TripNum	(6) UPC per trip	(7) ProductGroup per trip	(8) Dept per trip
NumSup	-0.203	0.0893	0.0659*	0.0317***				
	(0.211)	(0.0602)	(0.0349)	(0.00943)				
NumClubs					0.960***	-0.281***	-0.159***	-0.0364***
					(0.283)	(0.0664)	(0.0387)	(0.0112)
Observations	134,495	134,495	134,495	134,495	109,651	109,651	109,651	109,651
R-squared	0.755	0.778	0.790	0.766	0.766	0.789	0.799	0.771
Prob > F	0.336	0.138	0.059	0.001	0.001	0.000	0.000	0.001
Household-Quarter FE	X	X	X	X	X	X	X	X
Year-Quarter FE	X	X	X	X	X	X	X	X
Number of units	9371	9371	9371	9371	6814	6814	6814	6814
Number of clusters	9371	9371	9371	9371	6814	6814	6814	6814
BOTE	0.040	0.131	0.143	0.163	-0.126	-0.314	-0.279	-0.150

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Includes only households living in zip5 with store entry during the period. BOTE refers to a back-of-the-envelope estimate of how much each independent variable explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table F7: Difference between the most affected households and the least affected households for supercenter entry

	Mean Diff	sd	p-value
hh_hhi_yq	0.053	0.023	0.546
tripnum_yq	-1.931	2.204	1.000
numsup	0.038	0.040	1.000
numclubs	-0.082	0.050	0.887
ind.hhincomeQ1	**0.362**	0.177	0.680
ind.hhincomeQ2	**0.344**	0.169	0.682
ind.hhincomeQ3	-0.213	0.160	0.972
ind.hhincomeQ4	-0.494	0.162	0.152
ind.hhsize1	**0.283**	0.162	0.843
ind.hhsize2	-0.673	0.169	0.014**
ind.hhsize3plus	**0.390**	0.187	0.661
ind.ageAbove55	**-0.686**	0.210	0.101
ind.CentralMetro	-0.751	0.324	0.528
ind.LFringeMetro	0.126	0.270	1.000
ind.MediumMetro	-0.362	0.252	0.949
ind.SmallMetro	**0.719**	0.209	0.058*
ind.Micropolitan	0.105	0.115	1.000
ind.Noncore	0.164	0.082	0.710

Notes: The most affected group has top 10% sorted effect. The least affected group has bottom 10% sorted effect. The difference is average value for the most affected minus the least affected.

Table F8: Difference between the most affected households and the least affected households for clubs entry

	Mean Diff	sd	p-value
hh_hhi_yq	0.052	0.030	0.952
tripnum_yq	-7.433	2.746	0.495
num_sup	-0.176	0.116	0.983
numclubsoc	0.071	0.047	0.983
ind.hhincomeQ1	0.044	0.141	1.000
ind.hhincomeQ2	0.134	0.141	1.000
ind.hhincomeQ3	-0.552	0.175	0.319
ind.hhincomeQ4	0.374	0.228	0.962
ind.hhsize1	0.191	0.189	1.000
ind.hhsize2	-0.704	0.195	0.157
ind.hhsize3plus	0.514	0.222	0.732
ind.ageAbove55	-0.927	0.222	0.057*
ind.CentralMetro	-0.782	0.233	0.235
ind.LFringeMetro	0.849	0.306	0.480
ind.MediumMetro	-0.429	0.250	0.956
ind.SmallMetro	0.321	0.232	0.996
ind.Micropolitan	0.040	0.152	1.000
ind.Noncore	0.000	0.000	

Notes: The most affected group has top 10% sorted effect. The least affected group has bottom 10% sorted effect. The difference is average value for the most affected minus the least affected.

Table F9: Effect of Variety and Prices by Store Variety Rank, RMS

	(1)	(2)	(3)	(4)
Store variety rank	1-10	11-20	21-40	41+
VARIABLES	Household HHI			
Variety: UPCs	0.0775***	0.0296	0.0470***	-0.0804***
	(0.0109)	(0.0199)	(0.00495)	(0.00403)
Price Index	0.0373	-0.102**	0.0354**	-0.0928***
	(0.0237)	(0.0412)	(0.0155)	(0.0228)
Observations	630308	221091	769278	485836
R-squared	0.687	0.690	0.684	0.696
Prob > F	0.000	0.000	0.000	0.000
Number of units	49132	18776	72260	46359
Number of clusters	49132	18776	72260	46359
First stage F-stat	4299.009	8526.101	20949.600	21569.972
BOTE: All	0.219	0.078	0.191	-0.082
BOTE: Variety	0.216	0.092	0.187	-0.073
BOTE: Prices	0.003	-0.014	0.004	-0.009

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Store variety rank refers to the average number of UPCs per store in each chain, ranked from highest to lowest. Variety: UPCs refers to the number of UPCs per store, Variety: Modules refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 4.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table F10: Effect of Variety and Prices on Other Measures of Concentration, HMS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Num Retailers		Within-dept HHI		Cross-dept variance		Cross-HH variance	
Variety: UPCs	-0.204***		0.0159***		-0.00470***		-0.00131	
	(0.0789)		(0.00408)		(0.00174)		(0.00482)	
Variety: Modules		-0.939***		0.0551***		-0.0346***		-0.0106
		(0.322)		(0.0168)		(0.00733)		(0.0202)
RPI (US)	3.129***	3.005***	-0.194***	-0.189***	0.0323***	0.0265**	0.0217	0.0199
	(0.543)	(0.548)	(0.0284)	(0.0286)	(0.0125)	(0.0127)	(0.0324)	(0.0323)
Observations	3605864	3605864	3605864	3605864	3605864	3605864	3605864	3605864
R-squared	0.684	0.683	0.638	0.638	0.492	0.490	0.579	0.579
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000
Number of units	190795	190795	190795	190795	190795	190795	190795	190795
Number of clusters	190795	190795	190795	190795	190795	190795	190795	190795
First stage F-stat	428.909	72.872	428.909	72.872	428.909	72.872	428.909	72.872
BOTE: All	0.006	0.008	0.029	0.023	0.019	0.041	-0.001	-0.005
BOTE: Variety	0.013	0.015	0.049	0.043	0.025	0.046	-0.003	-0.006
BOTE: Prices	-0.006	-0.006	-0.020	-0.020	-0.006	-0.005	0.002	0.002

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety: UPCs refers to the number of UPCs per store, Variety: Modules refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 4.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table F11: Effect of Variety and Prices on Other Measures of Concentration, RMS

VARIABLES	(1) Num Retailers	(2)	(3) Within-dept HHI	(4)	(5) Cross-dept variance	(6)	(7) Cross-HH variance	(8)
Variety: UPCs	-1.058***		0.0864***		-0.00487		-0.0177	
	(0.312)		(0.0167)		(0.00713)		(0.0190)	
Variety: Modules		-3.083***		0.0143		-0.0476*		-0.201***
		(1.102)		(0.0584)		(0.0281)		(0.0742)
Price Index	0.675	0.596	0.00385	-0.0430	-0.0146	-0.0224	0.00145	-0.0334
	(0.607)	(0.636)	(0.0313)	(0.0328)	(0.0150)	(0.0160)	(0.0350)	(0.0377)
Observations	974443	974443	974443	974443	974443	974443	974443	974443
R-squared	0.718	0.718	0.681	0.681	0.522	0.522	0.631	0.631
Prob > F	0.000	0.000	0.000	0.146	0.990	0.004	0.998	0.000
Number of units	82135	82135	82135	82135	82135	82135	82135	82135
Number of clusters	82135	82135	82135	82135	82135	82135	82135	82135
First stage F-stat	1624.059	409.923	1624.059	409.923	1624.059	409.923	1624.059	409.923
BOTE: All	0.065	0.005	0.349	-0.006	0.023	0.012	-0.045	-0.023
BOTE: Variety	0.066	0.007	0.349	0.002	0.021	0.008	-0.046	-0.019
BOTE: Prices	-0.001	-0.002	0.000	-0.008	0.002	0.005	0.000	-0.004

Notes: Robust standard errors are in parentheses, clustered by household. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Variety: UPCs refers to the number of UPCs per store, Variety: Modules refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 4.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table F12: Effect of Variety and Prices on Trip Measures, HMS

VARIABLES	(1) TripNum	(2)	(3) UPC per trip	(4)	(5) Groups per trip	(6)	(7) Dept per trip	(8)
Variety: UPCs	-0.476		0.750***		0.407***		0.0571***	
	(0.404)		(0.0984)		(0.0586)		(0.00724)	
Variety: Modules		-1.664		3.171***		1.754***		0.271***
		(1.614)		(0.461)		(0.269)		(0.0357)
RPI (US)	-2.831	-3.002	-2.688***	-2.294***	-1.590***	-1.369***	-0.239***	-0.202***
	(2.819)	(2.816)	(0.691)	(0.726)	(0.408)	(0.425)	(0.0496)	(0.0533)
Observations	3605864	3605864	3605864	3605864	3605856	3605856	3605856	3605856
R-squared	0.726	0.726	0.767	0.762	0.779	0.776	0.695	0.687
Prob > F	0.102	0.321	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	190795	190795	190795	190795	190795	190795	190795	190795
Number of clusters	190795	190795	190795	190795	190795	190795	190795	190795
First stage F-stat	428.909	72.872	428.909	72.872	428.907	72.871	428.907	72.871
BOTE: All	0.009	0.008	0.077	0.084	0.067	0.075	0.326	0.406
BOTE: Variety	0.007	0.006	0.087	0.093	0.077	0.084	0.379	0.451
BOTE: Prices	0.001	0.002	-0.011	-0.009	-0.010	-0.009	-0.053	-0.045

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety: UPCs refers to the number of UPCs per store, Variety: Modules refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 4.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table F13: Effect of Variety and Prices on Trip Measures, RMS

VARIABLES	(1) TripNum	(2)	(3) UPC per trip	(4)	(5) Groups per trip	(6)	(7) Dept per trip	(8)
Variety: UPCs	-3.900***		1.702***		0.912***		0.150***	
	(1.472)		(0.430)		(0.237)		(0.0265)	
Variety: Modules		-0.782		2.062		0.738		0.264***
		(4.965)		(1.414)		(0.816)		(0.0980)
Price Index	-9.486***	-7.402**	1.674**	1.151	0.856**	0.493	0.0474	0.0197
	(3.083)	(3.183)	(0.743)	(0.793)	(0.433)	(0.460)	(0.0528)	(0.0558)
Observations	974443	974443	974443	974443	974443	974443	974443	974443
R-squared	0.768	0.768	0.795	0.795	0.807	0.807	0.722	0.722
Prob > F	0.000	0.000	0.000	0.034	0.000	0.941	0.000	0.000
Number of units	82135	82135	82135	82135	82135	82135	82135	82135
Number of clusters	82135	82135	82135	82135	82135	82135	82135	82135
First stage F-stat	1624.059	409.923	1624.059	409.923	1624.059	409.923	1624.059	409.923
BOTE: All	0.058	0.005	0.147	0.011	0.128	0.007	0.338	0.024
BOTE: Variety	0.055	0.000	0.144	0.006	0.125	0.004	0.335	0.022
BOTE: Prices	0.003	0.005	0.004	0.005	0.003	0.003	0.003	0.002

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. Variety: UPCs refers to the number of UPCs per store, Variety: Modules refers to the number of product modules per store, RPI (county) refers to the RPI using county-level reference prices as constructed in Section 4.1, RPI (US) uses national-level reference prices, and Price Index is a store price index constructed following Leung (2021). To calculate the BOTEs resulting from changes in the IV only, we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. For the store price index BOTE, we use the change in the RPI IV over the sample period, which reflects the degree to which larger chains changed prices relative to its competitors.

Table F14: Effect of shopping trips on household HHI, IV with region average wage and unemployment rate, county-time FE

	(1)	(2)	(3)
IV group	Age	Education	Both
VARIABLES		Household HHI	
Log Trips	-0.297***	-0.261***	-0.283***
	(0.0365)	(0.0613)	(0.0313)
Observations	2274654	2295327	2267467
R-squared	0.749	0.754	0.752
Prob > F	0.000	0.000	0.000
Number of units	127883	128376	127637
Number of clusters	127883	128376	127637
First stage F-stat	41.649	12.997	28.593
BOTE: Trips	1.546	1.245	1.506
BOTE: IVs	0.122	0.034	0.151

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. BOTE refers to a back-of-the-envelope estimate of how much each independent variable (log trips) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period. To calculate the BOTEs resulting from changes in the IV only (BOTE: IVs), we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively.

Table F15: Effect of shopping trips on other measures of concentration and trip measures, IV with household employment

VARIABLES	(1) Log Num Retailers	(2) Within-dept HHI	(3) Cross-dept variance	(4) Cross-HH variance	(5) UPC per trip	(6) Groups per trip	(7) Dept per trip
Log Trips	0.861***	-0.170***	0.0475***	-0.254***	-9.168***	-5.225***	-0.506***
	(0.0280)	(0.0127)	(0.00578)	(0.0151)	(0.196)	(0.113)	(0.0180)
Observations	3622992	3622992	3622992	3622992	3622992	3622990	3622990
R-squared	0.812	0.707	0.533	0.603	0.922	0.929	0.797
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	193527	193527	193527	193527	193527	193527	193527
Number of clusters	193527	193527	193527	193527	193527	193527	193527
First stage F-stat	116.170	116.170	116.170	116.170	116.170	116.180	116.180
BOTE: Trips	0.828	1.284	0.612	1.423	2.599	2.413	8.150

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. BOTE: Trips refers to a back-of-the-envelope estimate of how much each independent variable (log trips) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period.

Table F16: Effect of shopping trips on other measures of concentration and trip measures, IV with region average real wage and unemployment rate, age and education

VARIABLES	(1) Log Num Retailers	(2) Within-dept HHI	(3) Cross-dept variance	(4) Cross-HH variance	(5) UPC per trip	(6) Groups per trip	(7) Dept per trip
Log Trips	0.966***	-0.261***	0.0300**	-0.0613*	-10.16***	-5.616***	-0.695***
	(0.0675)	(0.0289)	(0.0132)	(0.0342)	(0.499)	(0.275)	(0.0444)
Observations	2278148	2278148	2278148	2278148	2278148	2278147	2278147
R-squared	0.815	0.708	0.544	0.625	0.926	0.933	0.795
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of units	128038	128038	128038	128038	128038	128038	128038
Number of clusters	128038	128038	128038	128038	128038	128038	128038
First stage F-stat	30.737	30.737	30.737	30.737	30.737	30.730	30.730
BOTE: Trips	1.031	2.182	0.429	0.381	3.200	2.880	12.440
BOTE: IVs	0.096	0.203	0.040	0.035	0.297	0.267	1.155

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. BOTE: Trips refers to a back-of-the-envelope estimate of how much each independent variable (log trips) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period. To calculate the BOTEs resulting from changes in the IV only (BOTE: IVs), we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively.

Table F17: Effect of shopping trips on household HHI, IV with state and city gas prices

	(1)	(2)	(3)
	EIA: State	EIA: City	BLS: City
VARIABLES	Household HHI		
Log Trips	-0.414**	-0.485*	-0.457**
	(0.164)	(0.252)	(0.195)
Observations	719840	426514	584339
R-squared	0.703	0.658	0.678
Prob > F	0.000	0.000	0.000
Number of units	49720	29902	41775
Number of clusters	49720	29902	41775
First stage F-stat	8.128	3.993	6.089
BOTE: Trips, 2004-2019	1.969	2.303	2.169
BOTE: IVs, 2004-2019	0.108	0.116	0.129
BOTE: Trips, 2010-2016	0.595	0.696	0.656
BOTE: IVs, 2010-2016	-0.068	-0.062	-0.079

Notes: Robust standard errors are in parentheses, clustered by household. *** p<0.01, ** p<0.05, * p<0.1. BOTE: Trips refers to a back-of-the-envelope estimate of how much each independent variable (log trips) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period. To calculate the BOTEs resulting from changes in the IV only (BOTE: IVs), we multiply the change in the IVs by their respective first-stage coefficients for each independent variable, and then further multiply the changes in each variable by their second-stage coefficients respectively. We split these BOTEs into those for changes from 2004-2019 and 2010-2016 respectively.

Table F18: Effect of Online Shopping (Upper Bound)

VARIABLES	(1) HH Offline HHI	(2)	(3) HH HHI	(4)	(5) Offline Trip Number	(6) Offline Number of Retailers
Online Share	0.0578***	0.0565***	-0.174***	-0.176***	-14.51***	-1.774***
	(0.00582)	(0.00576)	(0.00637)	(0.00626)	(0.486)	(0.0701)
log(income)	0.00285***	0.00285***	0.00244***	0.00260***	-0.576***	-0.0287**
	(0.000657)	(0.000459)	(0.000607)	(0.000441)	(0.0680)	(0.0133)
Household Size	0.000953*	0.000357	0.000871*	0.000303	0.736***	0.0609***
	(0.000516)	(0.000380)	(0.000493)	(0.000368)	(0.0523)	(0.0110)
Observations	1432259	2409750	1433160	2410651	1432259	1432259
R-squared	0.687	0.711	0.686	0.712	0.761	0.720
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000
Number of clusters	49	49	49	49	49	49
Year-quarter FE	X	X	X	X	X	X
HH FE	X	X	X	X	X	X
HH-quarter FE	X	X	X	X	X	X
BOTE	0.013	0.012	-0.046	-0.046	0.026	0.014

Notes: Robust standard errors are in parentheses, clustered by state. *** p<0.01, ** p<0.05, * p<0.1. Column (2) and (4) use all households and the rest of columns use households who have reported online shopping at least once. BOTE refers to a back-of-the-envelope estimate of how much the independent variable (online share) explains the rise in household retail concentration. This is calculated by multiplying the estimated coefficient by the total change in the independent variable over the sample period, then dividing this number by the total change in household retail concentration in the sample period. Household online share increased by 0.012 from 2004 to 2016. Household offline HHI increased by 0.055, HH HHI 0.0458. Offline trip number decreased by 6.62 per quarter, offline number of retailers 1.54 per quarter.

Table F19: Effect of Amazon tax on household concentration

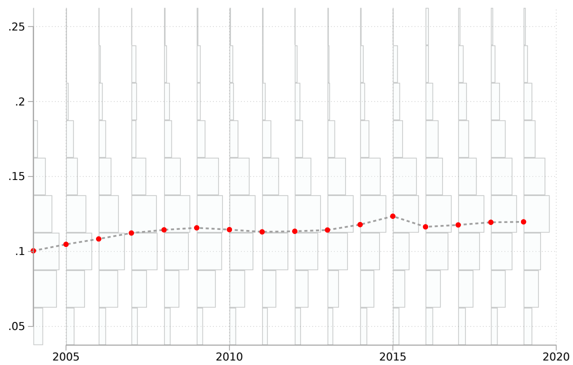
VARIABLES	(1) Household Retail Concentration	(2)	(3)	(4) Number of Trips	(5)	(6)
$TreatedStates \times I(t \geq treatedYQ)$	-0.0006			0.2155***		
	(0.0008)			(0.0693)		
$TreatedStates \times I(t = treatedYQ - 1)$		0.0001			-0.0790	
		(0.0008)			(0.0703)	
$TreatedStates \times I(t = treatedYQ)$		-0.0007			0.0306	
		(0.0009)			(0.0606)	
$TreatedStates \times I(t = treatedYQ + 1)$		0.0012			-0.0827	
		(0.0009)			(0.0720)	
$TreatedStates \times I(t \geq treatedYQ) \times TaxRate$			0.003			2.8245**
			(0.0120)			(1.1970)
Observations	2357822	2357822	2357822	2357822	2357822	2357822
R-squared	0.6687	0.6687	0.6687	0.7367	0.7367	0.7367
Prob > F	0.001	0.0049	0.0018	0.000	0.000	0.000
Number of clusters	2927	2927	2927	2927	2927	2927
Fixed Effects	X	X	X	X	X	X
Household Income	X	X	X	X	X	X

Notes: Robust standard errors are in parentheses, clustered by county. *** p<0.01, ** p<0.05, * p<0.1. From 2006 to 2016, 21 states are treated. The treatment has no effect on household online share, household offline concentration or overal household concentration.

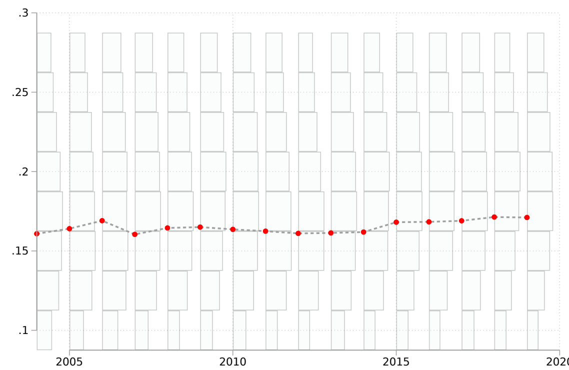
G Figures

Figure G1: Retail Concentration Over Time

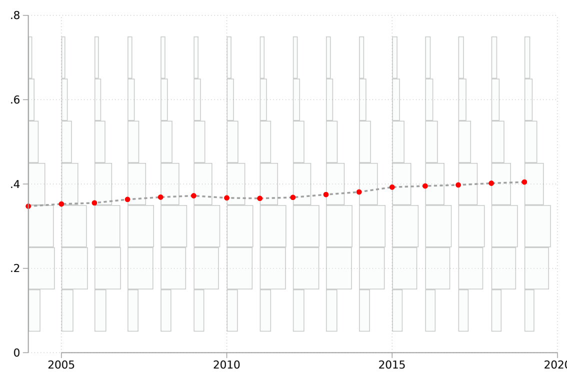
(a) Market



(b) County



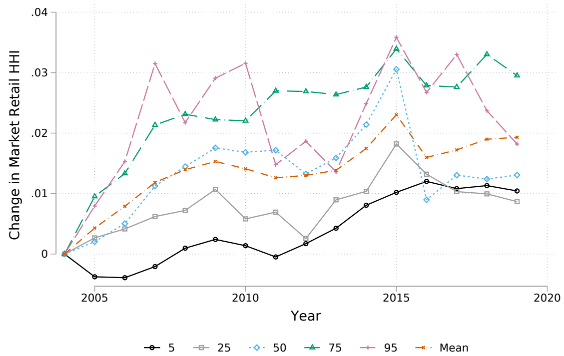
(c) Household



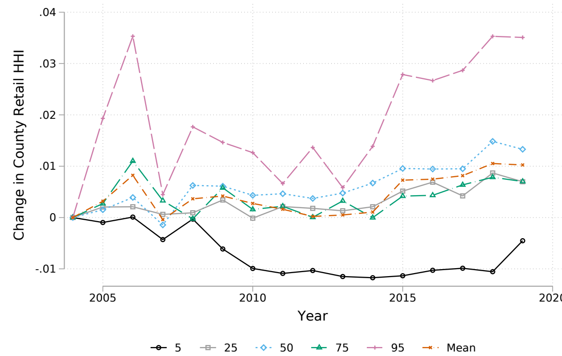
Notes: This figure plots the revenue-weighted average and distribution of retail concentration at the market, county, and household level respectively.

Figure G2: Retail Concentration Over Time By Percentile

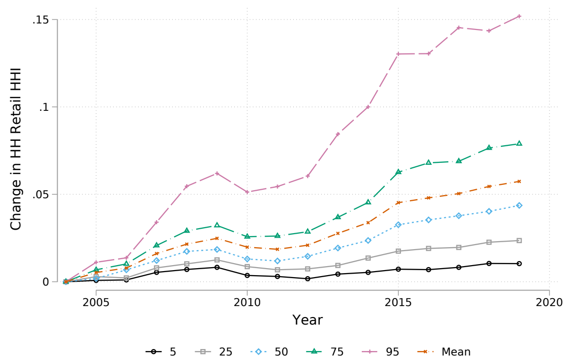
(a) Market



(b) County



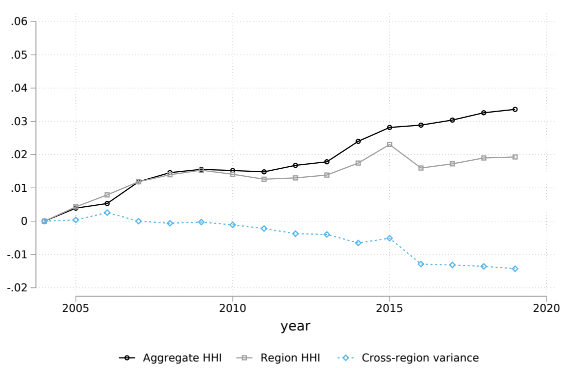
(c) Household



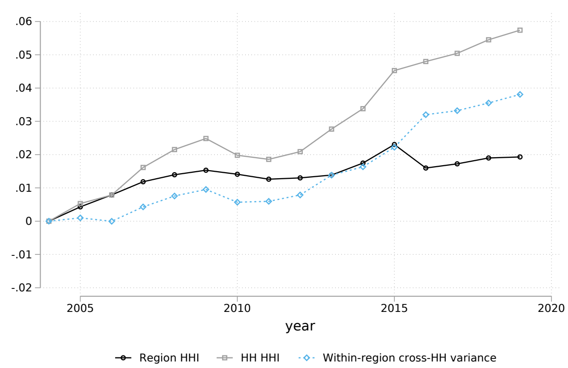
Notes: This figure plots the revenue-weighted changes of of retail concentration at the market, county, and household level respectively by percentile.

Figure G3: Decomposing Changes in Retail Concentration

(a) National and Market



(b) Market and Household

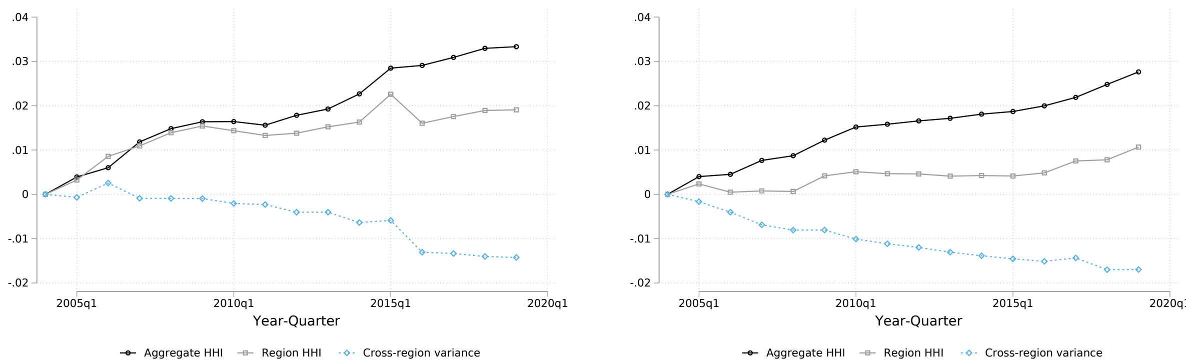


Notes: This figure shows that HHI can be decomposed at each level following Radaelli and Zenga (2002). For each term, we plot the yearly change over the sample period.

Figure G4: National and Market Concentration

(a) Revenue

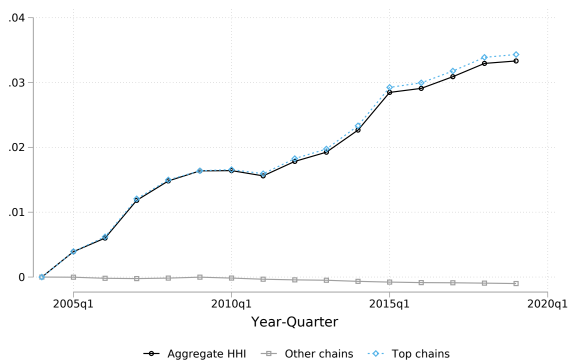
(b) Number of Stores



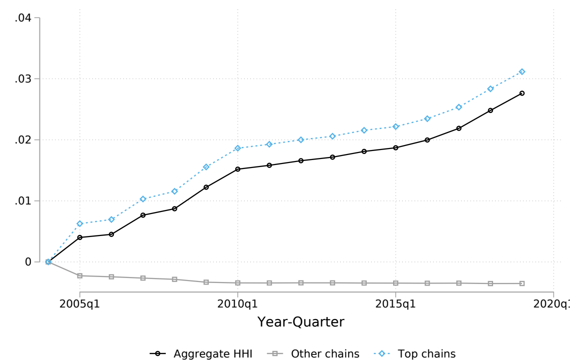
Notes: This figure shows that HHI can be decomposed at each level following Radaelli and Zenga (2002). For each term, we plot the yearly change over the sample period. We show this decomposition using two outcomes: revenue and the number of stores respectively.

Figure G5: Decomposing Changes in Retail Concentration by Groups of Chains

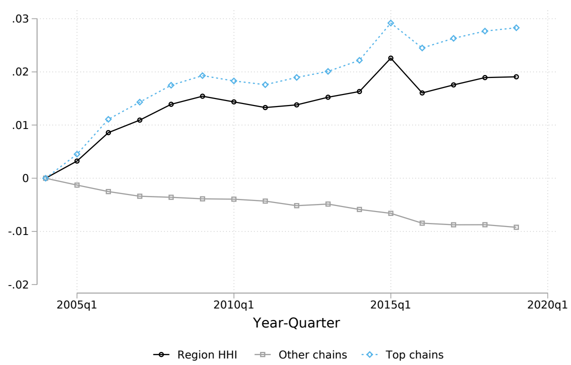
(a) Aggregate HHI, revenue



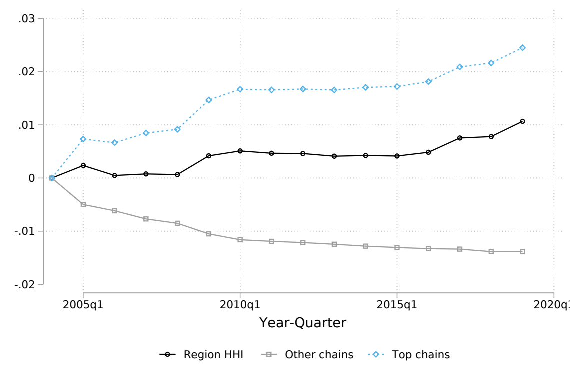
(b) Aggregate HHI, number of stores



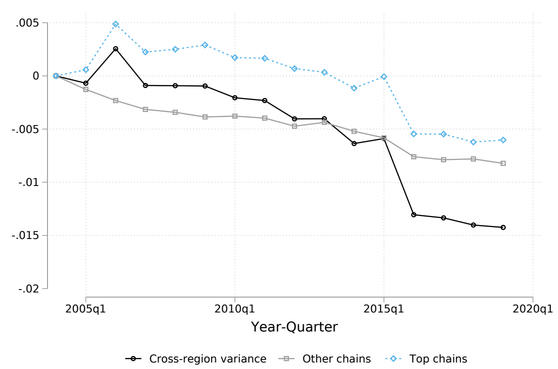
(c) Market HHI, revenue



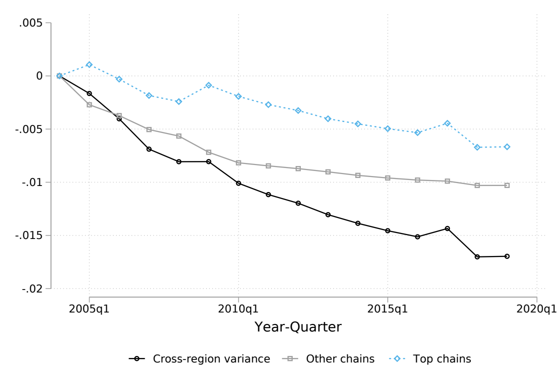
(d) Market HHI, number of stores



(e) Cross-market variance, revenue



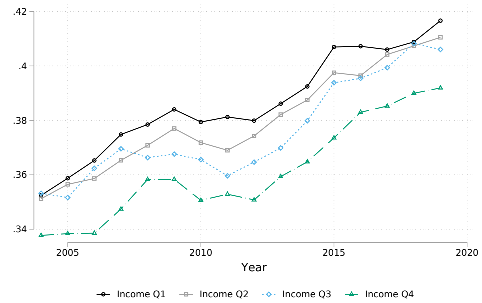
(f) Cross-market variance, number of stores



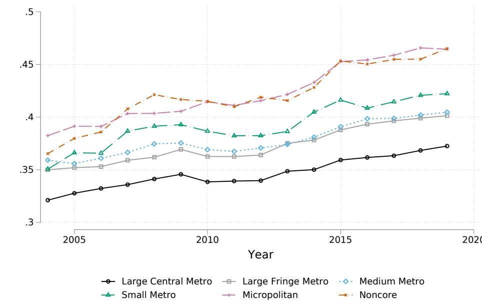
Notes: This figure shows that HHI can be decomposed at the market level following Radaelli and Zenga (2002) and further decomposed into the contributions by different groups of retail chains. For each term, we plot the yearly change over the sample period. We show this decomposition using two outcomes: revenue and the number of stores respectively. We group retail chains into two groups: the top 40 retail chains by revenue rank over the sample period, and other chains outside the top 40.

Figure G6: Household Retail Concentration by Demographic Group

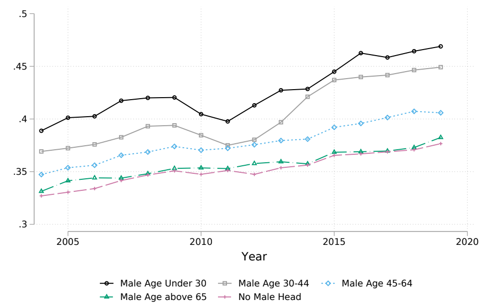
(a) Household Income



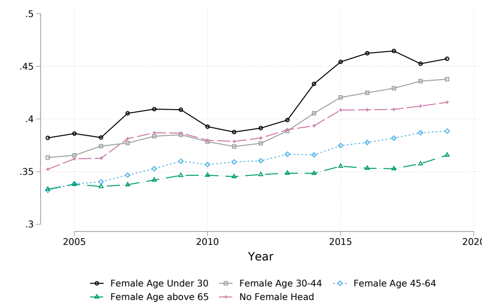
(b) Urban/Rural Counties



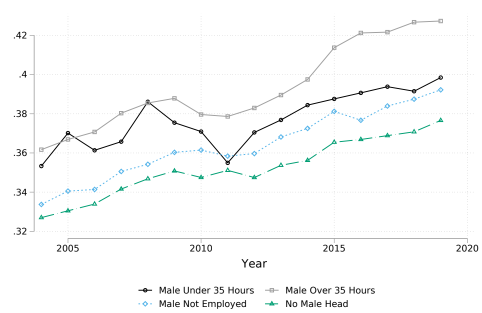
(c) Male Head Age



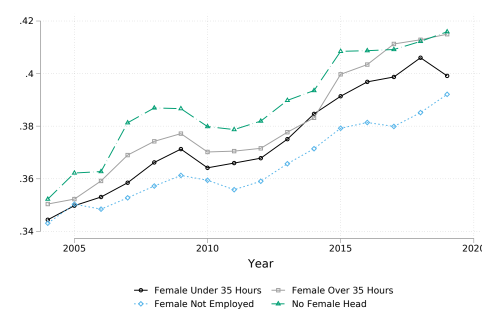
(d) Female Head Age



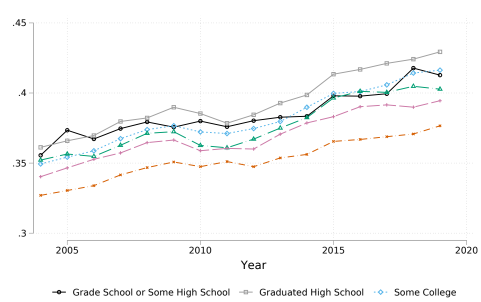
(e) Male Head Employment



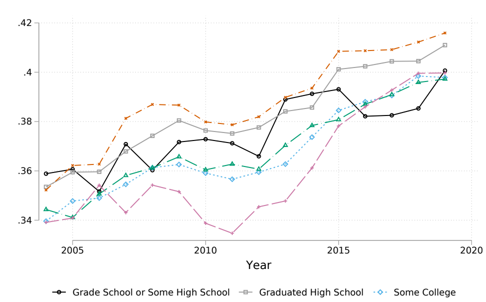
(f) Female Head Employment



(g) Male Head Education



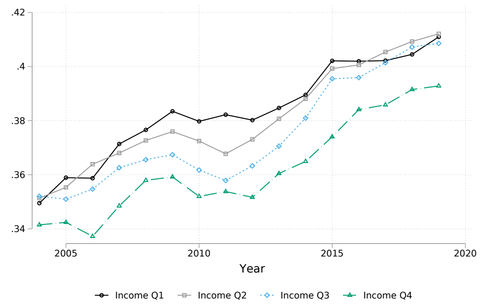
(h) Female Head Education



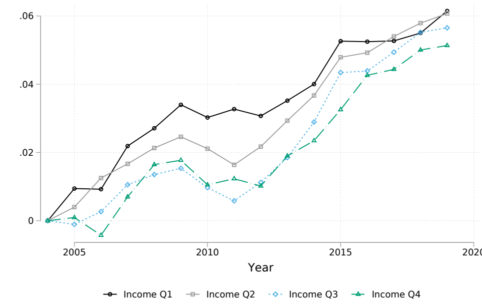
Notes: This figure plots the revenue-weighted average of retail concentration at the household level for different demographic groups. Panel (a) classifies households into four household income quartiles after adjusting for household size, with Q1 being the lowest income quartile and Q4 the highest income quartile. Panel (b) classifies households by the county they reside in and groups them based on the urban/rural classification of the National Center for Health Statistics (NCHS). Panel (c) and (d) classifies households by their male and female head's age. Panel (e) and (f) classifies households by their male and female head's hours of work. Panel (g) and (h) classifies households by their male and female head's level of education.

Figure G7: Household Retail Concentration by Demographic Group

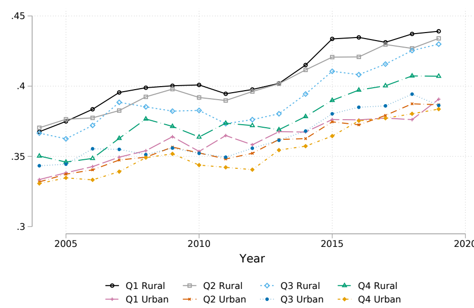
(a) Household Income, No Size Adjustment, Levels



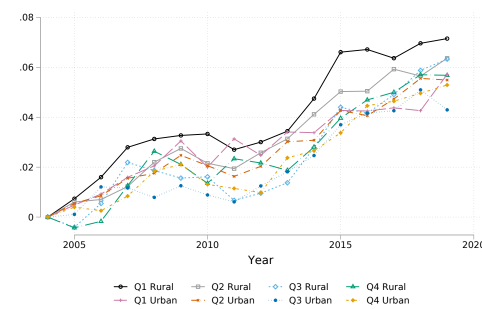
(b) Household Income, No Size Adjustment, Changes



(c) Household Income and Urban/Rural, Levels



(d) Household Income and Urban/Rural, Changes



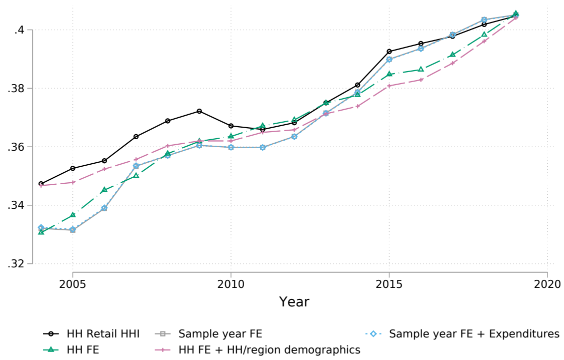
(e) Household Size, Levels

(f) Household Size, Changes

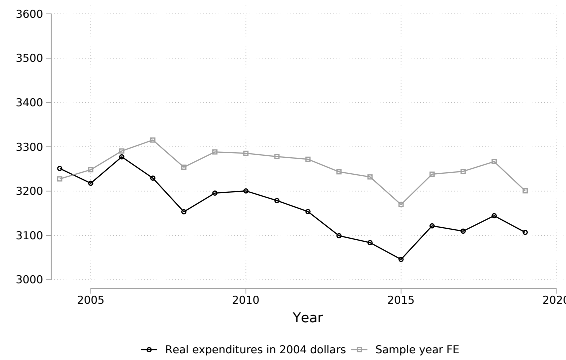
Notes: This figure plots the levels and changes (since 2004) of revenue-weighted average of retail concentration at the household level for different demographic groups. Panel (a) and (b) classify households into four household income quartiles, with Q1 being the lowest income quartile and Q4 the highest income quartile. Panel (c) and (d) classify households by the county they reside in and groups them based on the urban/rural classification of the National Center for Health Statistics (NCHS) and interact with household income after adjusting for household size. Panel (e) and (f) classify households by their household size.

Figure G8: Non-Magnet Products Only

(a) Household HHI



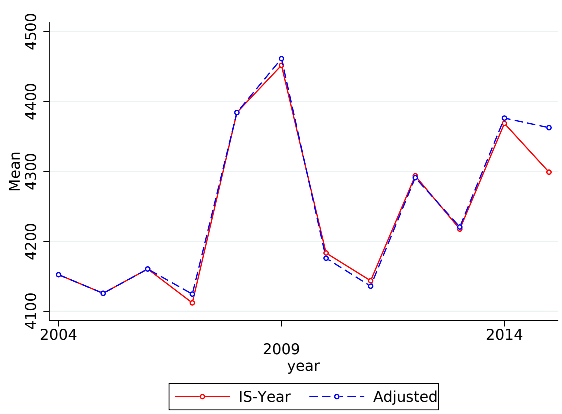
(b) Real Expenditures



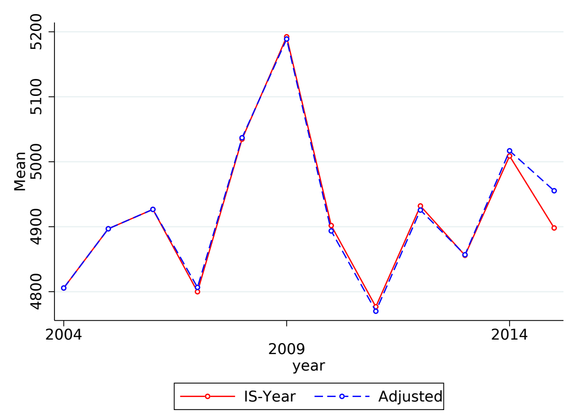
Notes: This figure shows trends for average household HHI and annual real expenditures from 2004-2019.

Figure G9: Real expenditures, CEX

(a) Food at home



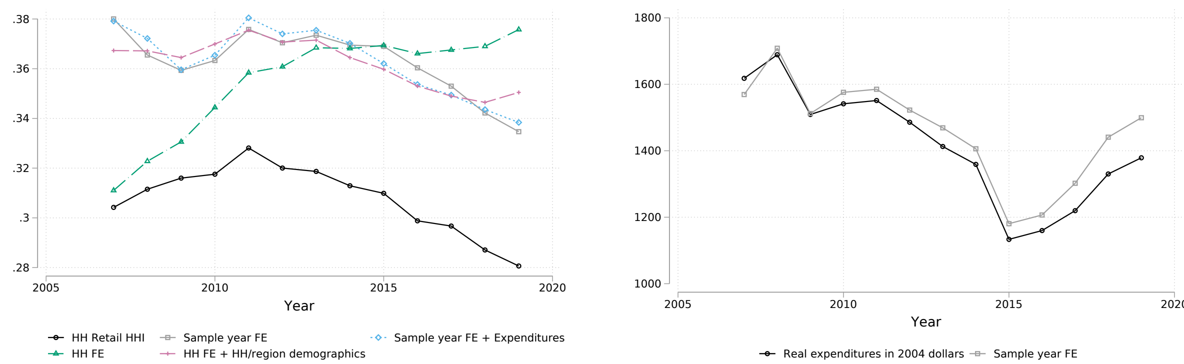
(b) Nielsen non-durable goods



Notes: This figure shows the average annual expenditure of food-at-home and matched non-durable goods captured by Nielsen, as classified by Coibion et al. (2021), in the Consumption Expenditure Survey (CEX) Interview Survey (IS). The adjusted series follows Coibion et al. (2021) in controlling for various household demographics.

Figure G10: Magnet Products Only

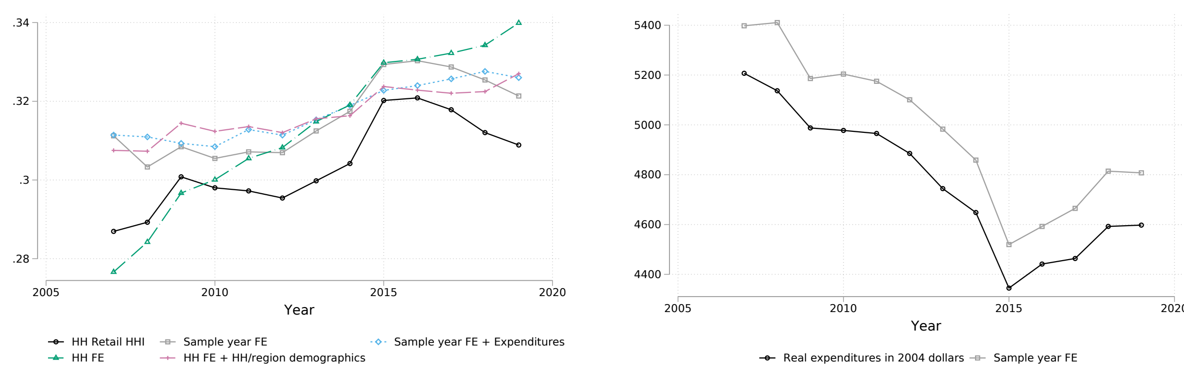
(a) Household HHI (b) Real Expenditures



Notes: This figure shows trends for average household HHI and annual real expenditures from 2007-2019.

Figure G11: All products, both magnet and non-magnet

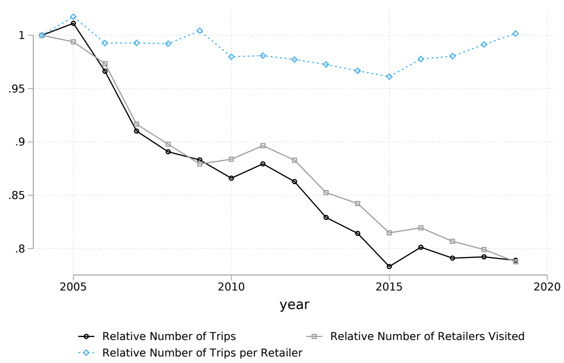
(a) Household HHI (b) Real Expenditures



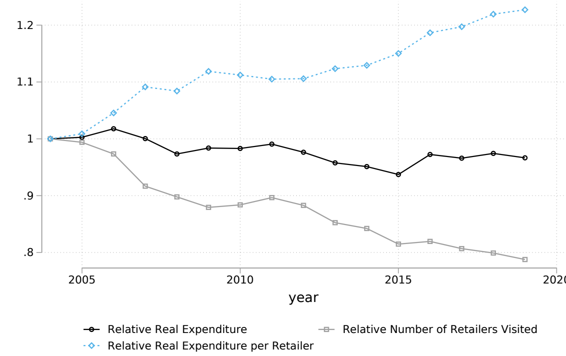
Notes: This figure shows trends for average household HHI and annual real expenditures from 2007-2019.

Figure G12: Decomposing trips, expenditures, and number of retailers per household

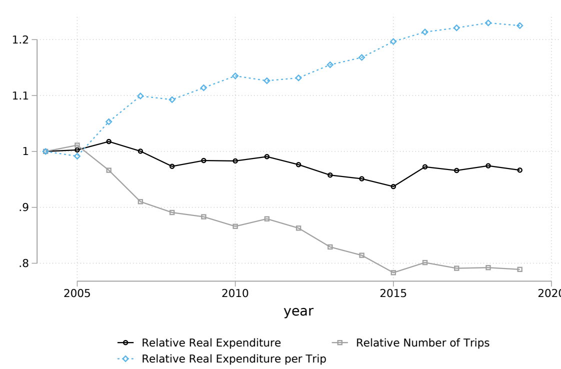
(a) Number of trips, number of retailers, and trips per retailer



(b) Real expenditure, number of retailers, and expenditures per retailer



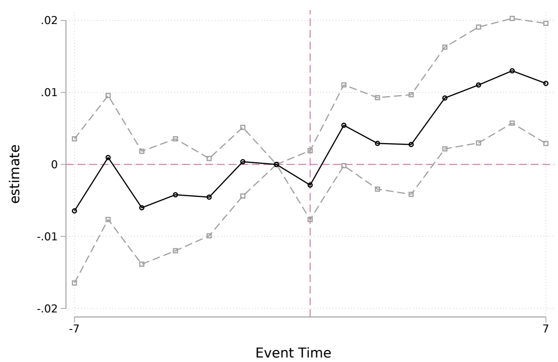
(c) Real expenditure, number of trips, and expenditures per trip



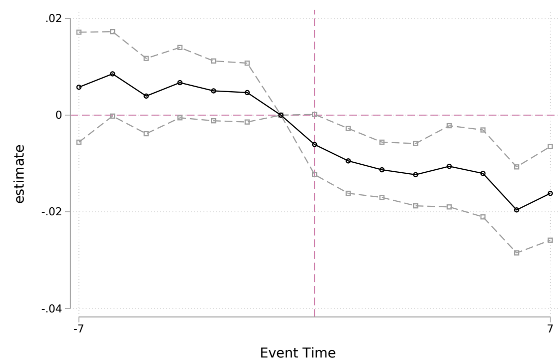
Notes: In Panel (a), this figure plots the annual sampling-weighted average of households' number of shopping trips per quarter, number of retailers visited per year, and trips per retailer, divided by their level in 2004. In Panel (b), this figure plots the annual sampling-weighted average of households' real expenditure, number of retailers visited per year, and expenditures per retailer, divided by their level in 2004. In Panel (c), this figure plots the annual sampling-weighted average of households' real expenditure, number of trips, and expenditures per trip, divided by their level in 2004.

Figure G13: Event Study Graph for Store Entry

(a) Supercenter Entry



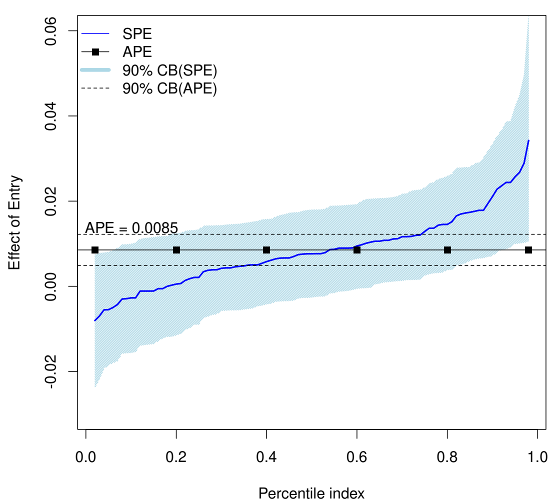
(b) Club Entry



Notes: This figure plots the sum of estimated coefficients for each period, along with the 95% confidence intervals, from regressions using a distributed lag model following de Chaisemartin and D'Haultfœuille (2020), where household retail HHI is regressed on the number of supercenters and club stores respectively. Household and period fixed effects are included.

Figure G14: Sorted Effects across Households

(a) Supercenter Entry



(b) Club Entry

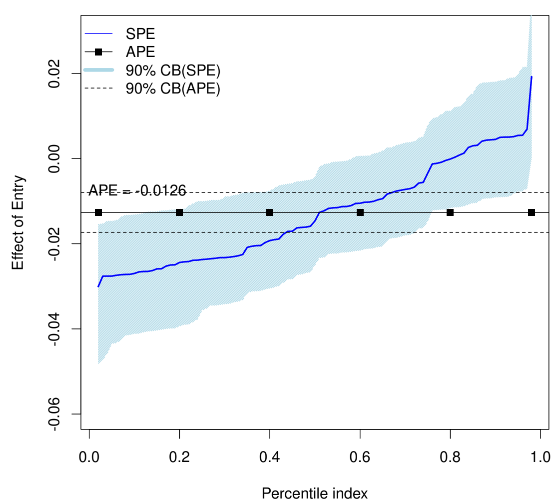
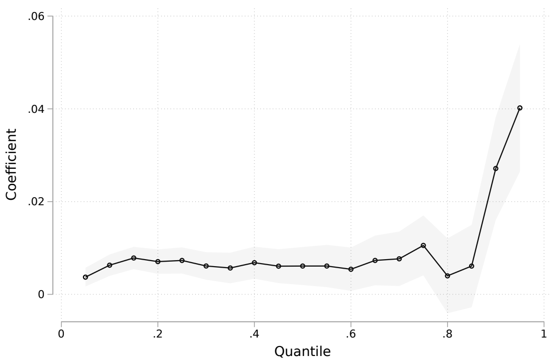


Figure G15: Quantile Treatment Effects

(a) Supercenter Entry



(b) Club Entry

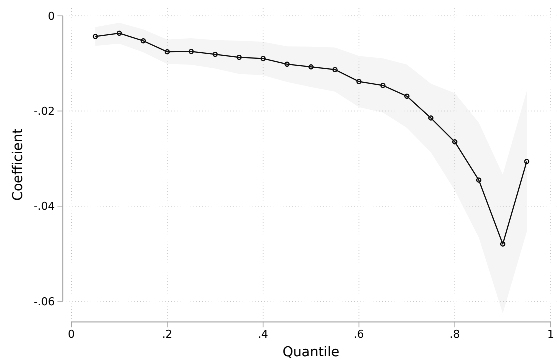
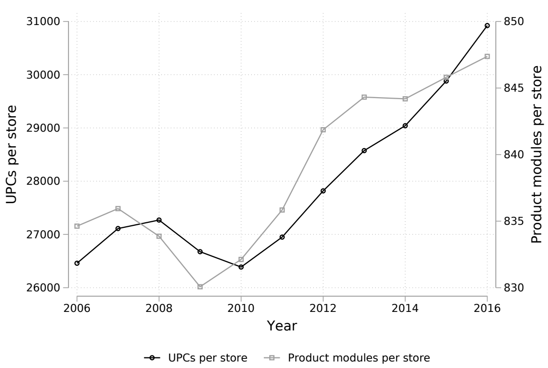
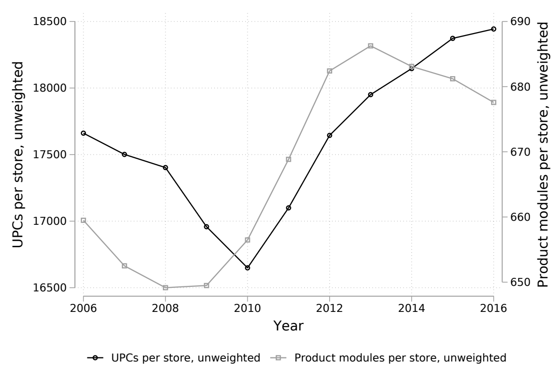


Figure G16: UPCs and product modules per store, RMS

(a) Store revenue-weighted



(b) Unweighted



Notes: This figure shows the number of UPCs and product modules per store in the RMS. The average is either calculated as the weighted average across all stores using store revenue as weights or is unweighted.

Figure G17: Employment Rates for Females Aged 25-54, 2004-2019

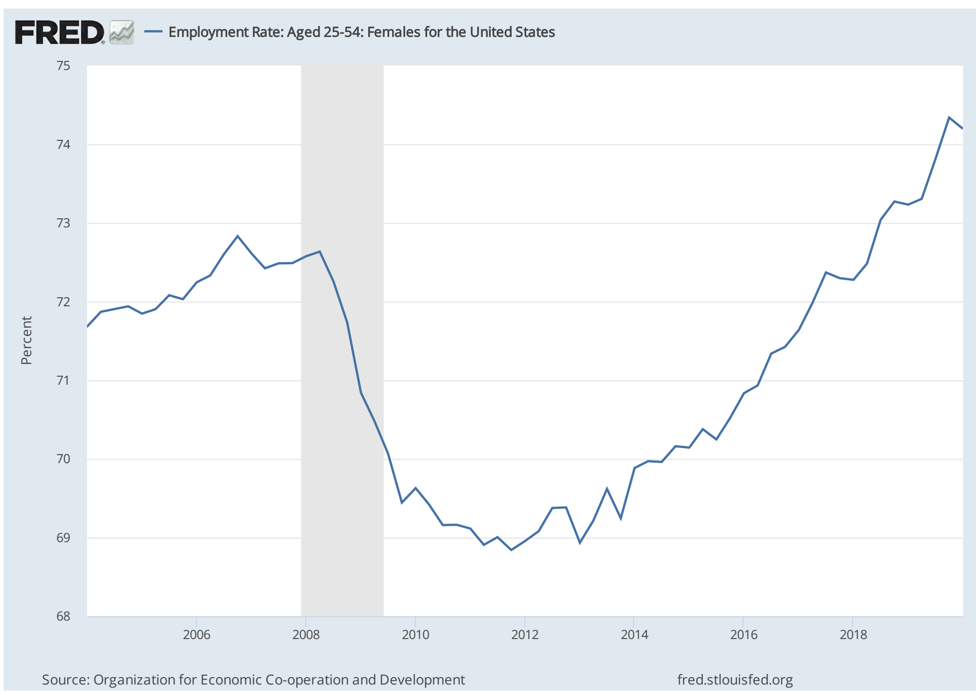
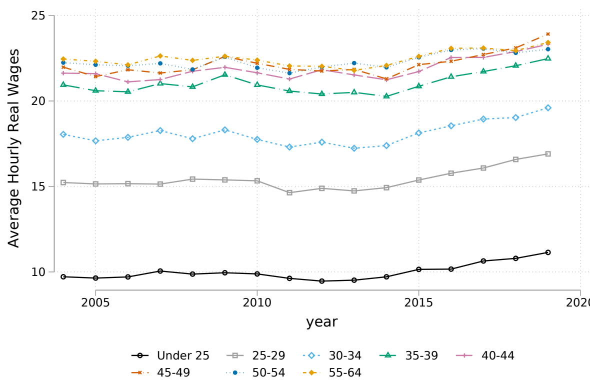
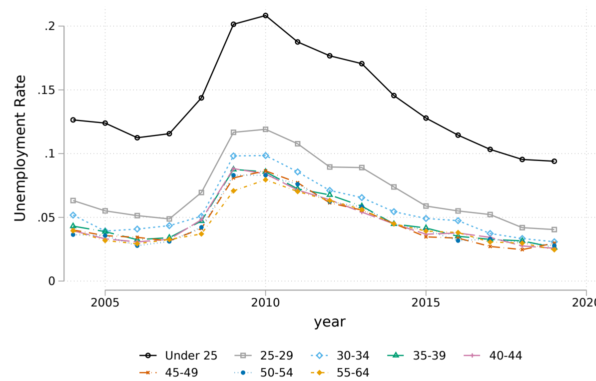


Figure G18: Time Costs by Age Group

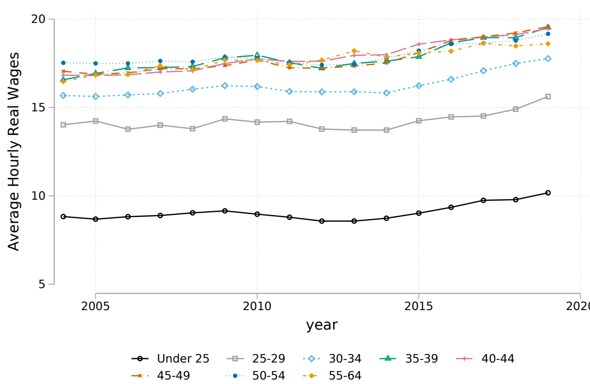
(a) Average Real Wages, Men



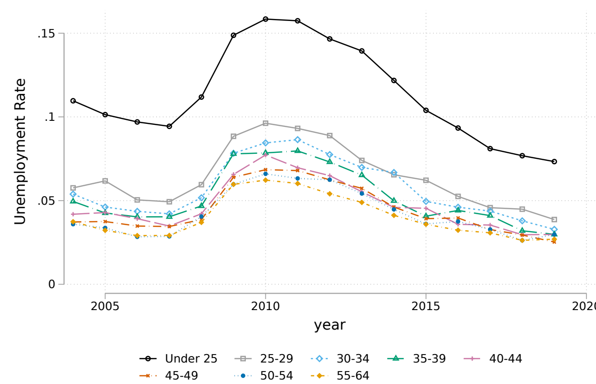
(b) Unemployment Rate, Men



(c) Average Real Wages, Women



(d) Unemployment Rate, Women

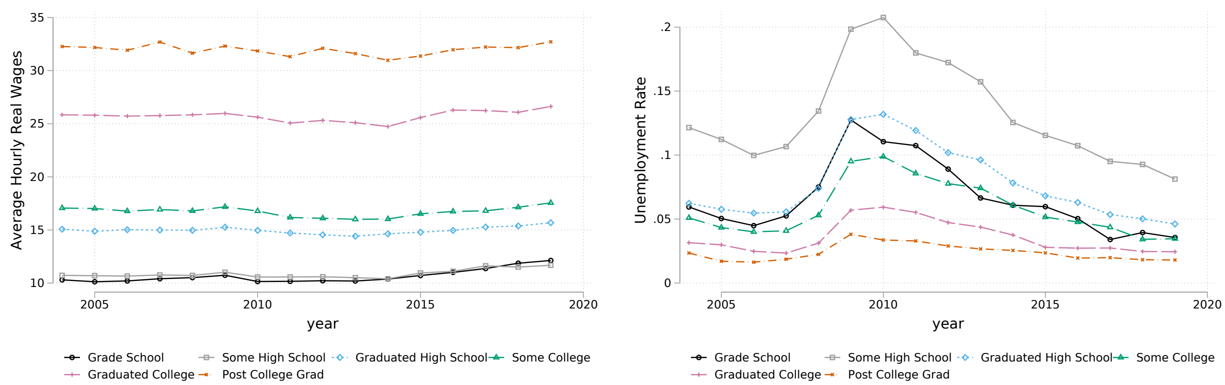


Notes: This figure plots average hourly real wages and unemployment rate by gender and age group based on the Current Population Survey Merged Outgoing Rotation Groups (CPS MORG).

Figure G19: Time Costs by Education

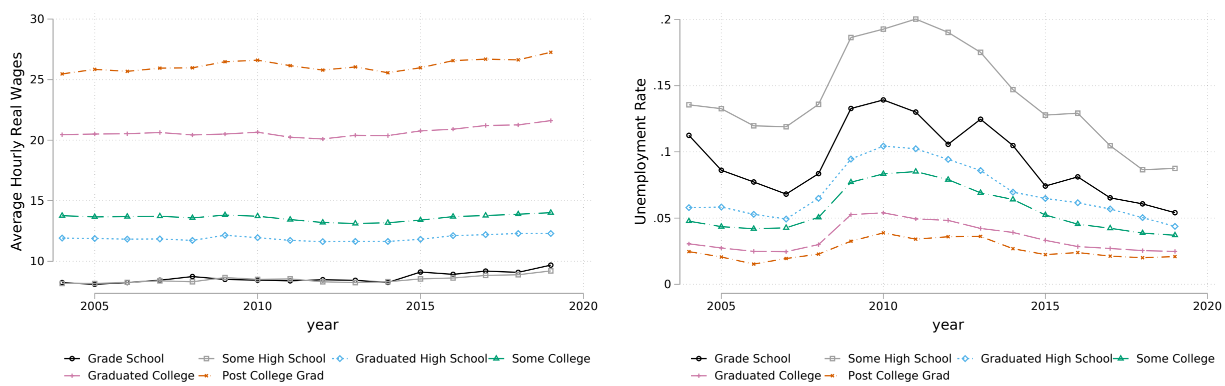
(a) Average Real Wages, Men

(b) Unemployment Rate, Men



(c) Average Real Wages, Women

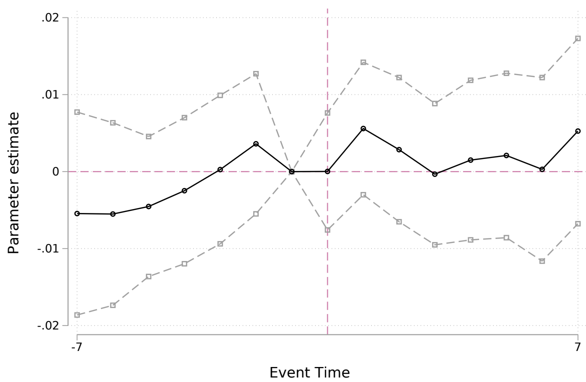
(d) Unemployment Rate, Women



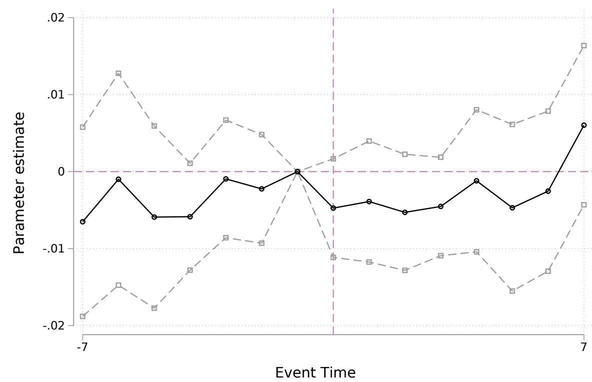
Notes: This figure plots average real hourly wages and unemployment rate by gender and education group based on the Current Population Survey Merged Outgoing Rotation Groups (CPS MORG).

Figure G20: Effects of non-employment, male heads

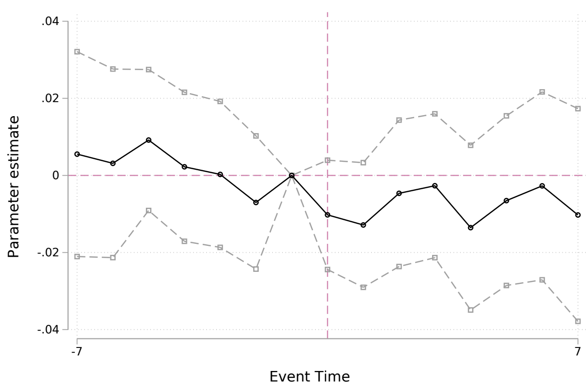
(a) Household HHI, OLS



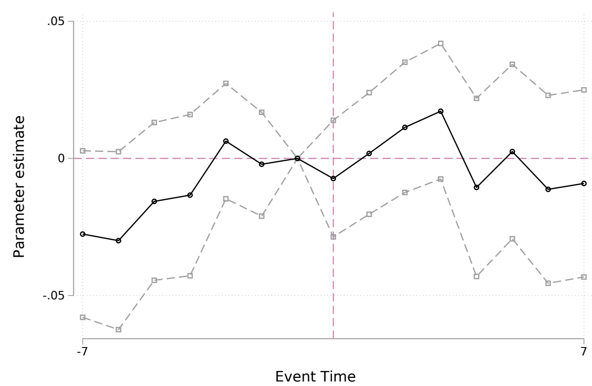
(b) Household HHI, DCDH



(c) Log trips, OLS



(d) Log trips, DCDH



Notes: This figure plots the sum of estimated coefficients for each period, along with the 95% confidence intervals, from regressions using a distributed lag model, where household retail HHI is regressed on an indicator for whether the household is not working. The sample focuses on prime-aged household heads from age 25-54. Household and period fixed effects are included. We use both OLS and methods robust to heterogeneous treatment effects following de Chaisemartin and D'Haultfœuille (2020).

Figure G21: US Regular All Formulations Gas Price, 2004-2019

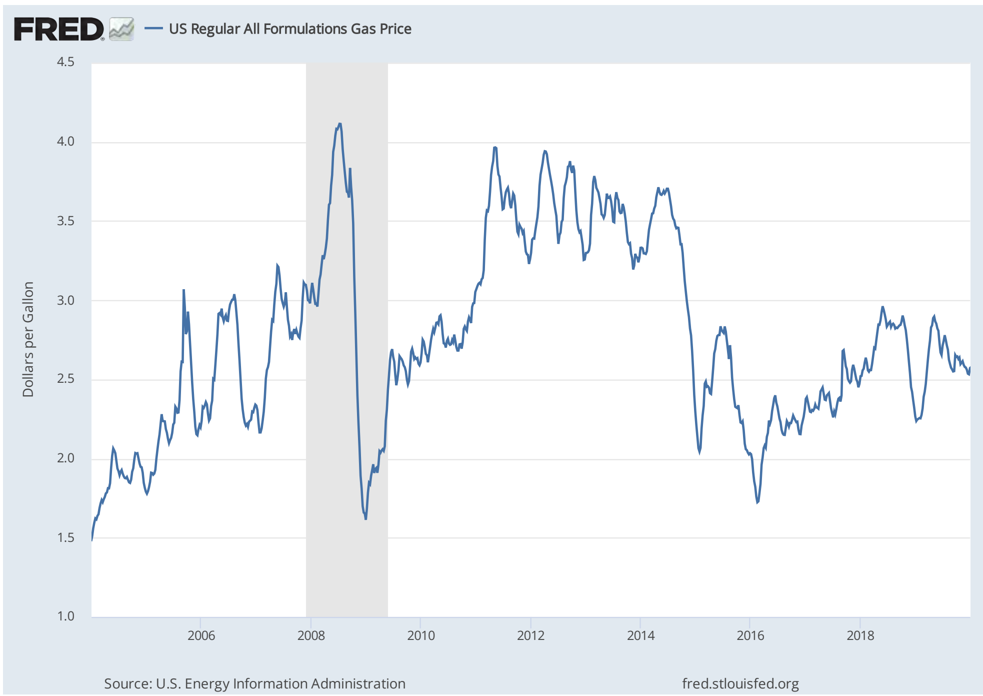


Figure G22: Household Concentration and Online Shopping

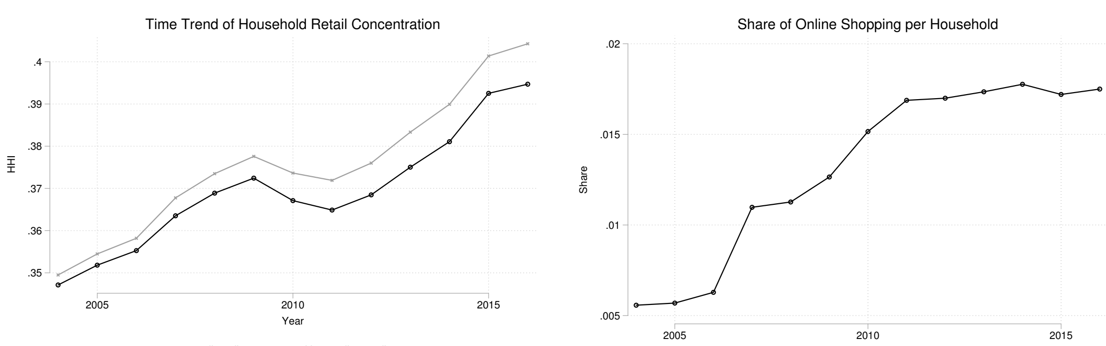


Figure G23: Share of Online Shopping in Each Department

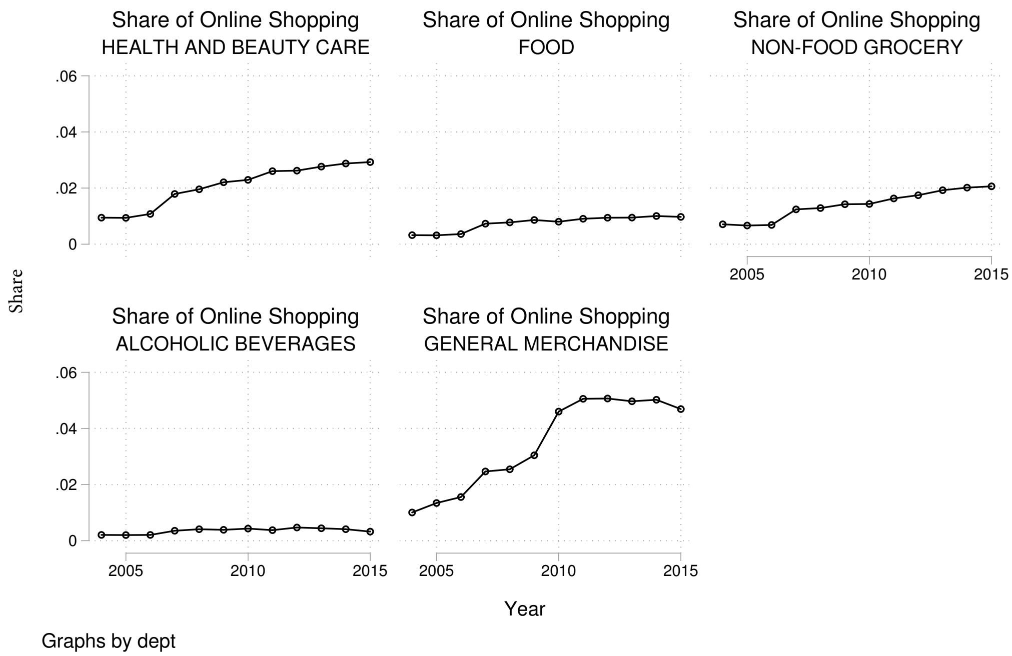


Figure G24: Household Concentration With and Without Online Shopping

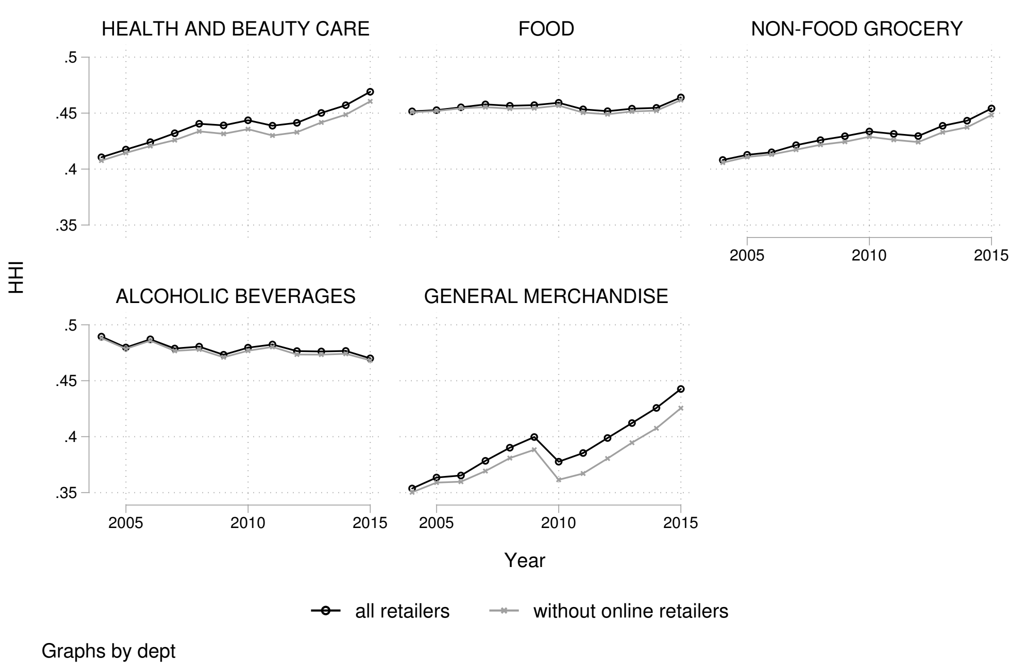
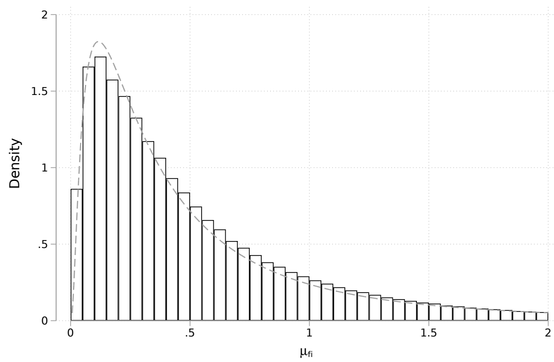
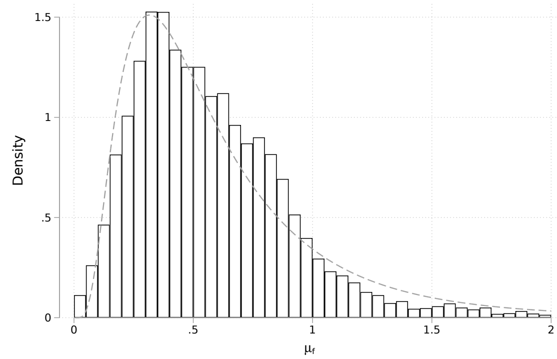


Figure G25: Distribution of shopping costs, upper bounds

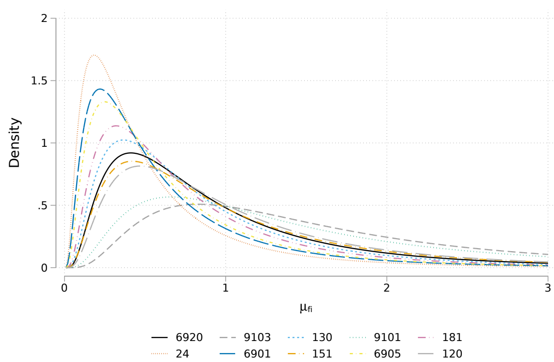
(a) Household-firm-year



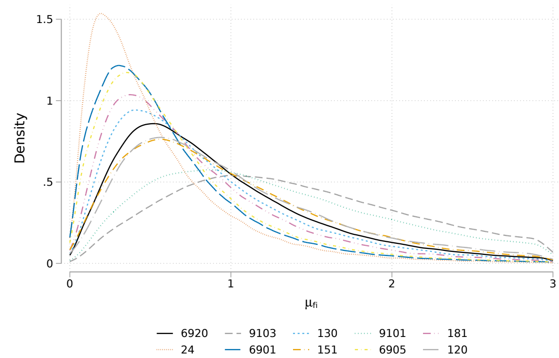
(b) Firm-year averages



(c) Household-year by firm, log-normal fit, top 10



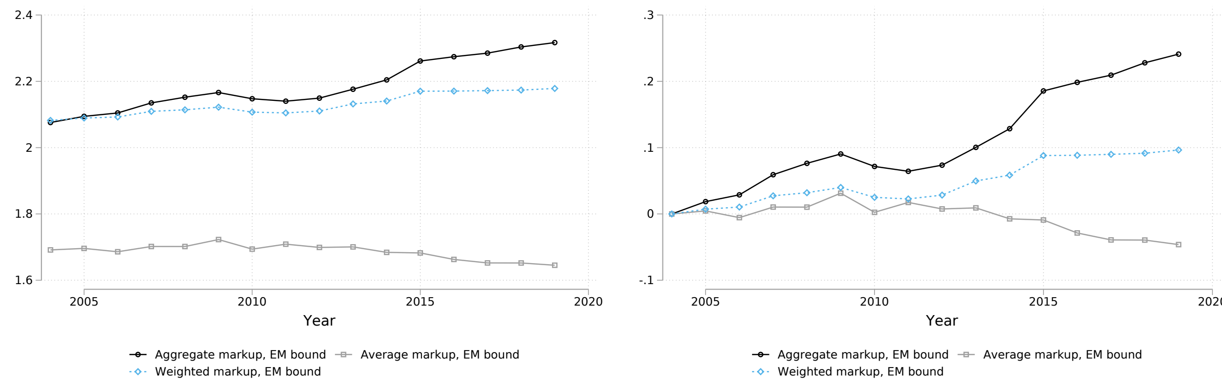
(d) Household-year by firm, kernel density fit, top 10



Notes: This figure shows the distribution of the upper bound of shopping costs μ_{fi} derived from equation (22) at different levels. In Panel (a), we plot the entire distribution of household-firm-year shopping costs. In Panel (b), we plot the averages at the firm-year level. In Panel (c), we plot the household-year distribution by firms for the top 10 firms, fitted to a log-normal distribution. In Panel (d), we plot the household-year distribution by firms for the top 10 firms, fitted to a kernel density function.

Figure G26: Aggregate, average, and weighted Cournot markups, 2004-2019

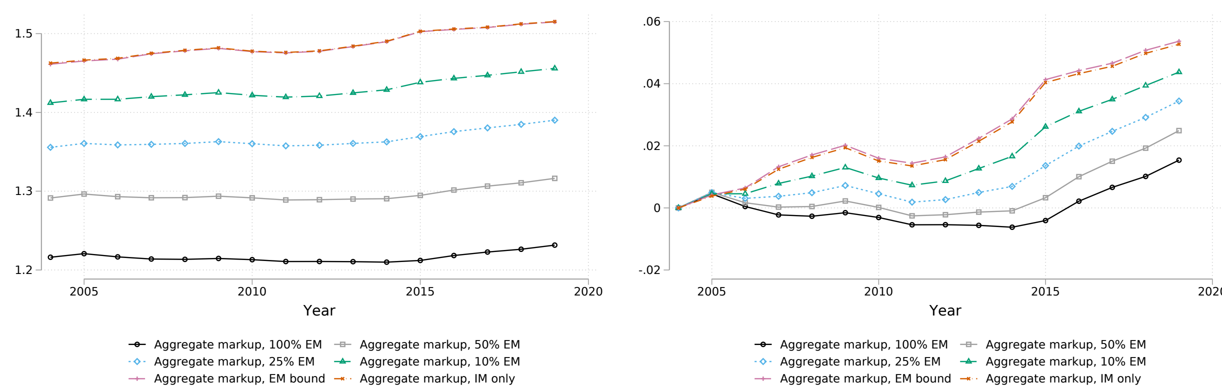
(a) Levels (b) Changes since 2004



Notes: This figure plots the trends in firm markups from 2004-2019 when aggregated using different weights, assuming Cournot competition, both in levels and the magnitude of changes since 2004. The default aggregate markup uses firm market shares over time. Average markup uses a simple average for firms that were present in the entire sample period. Weighted markup uses a weighted average, where weights are fixed to the aggregate market share across the entire sample period, for firms that were present in the entire sample period.

Figure G27: Aggregate markups using different extensive margin bounds, 2004-2019

(a) Levels (b) Changes since 2004

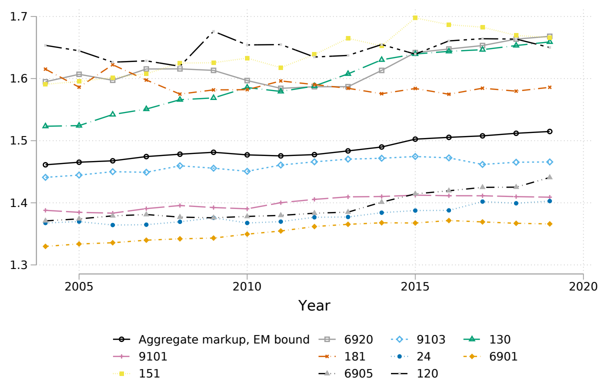


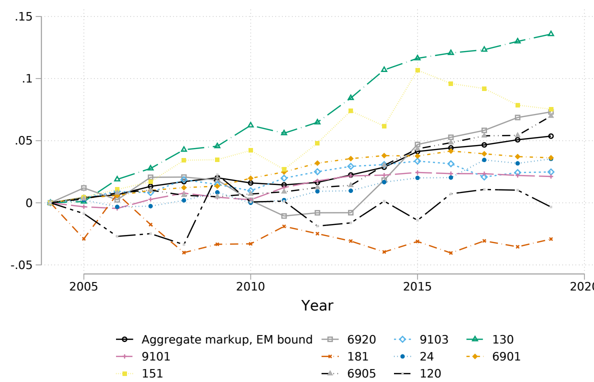
Notes: This figure plots the trends in firm markups from 2004-2019 when aggregated using different weights, assuming Bertrand competition, both in levels and the magnitude of changes since 2004.

Figure G28: Firm markups, top 10 firms, 2004-2019

(a) Levels

(b) Changes since 2004



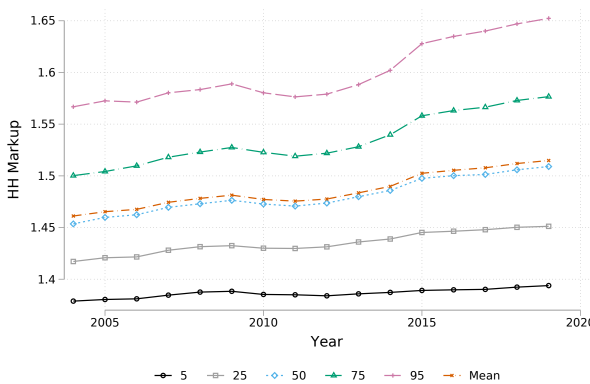


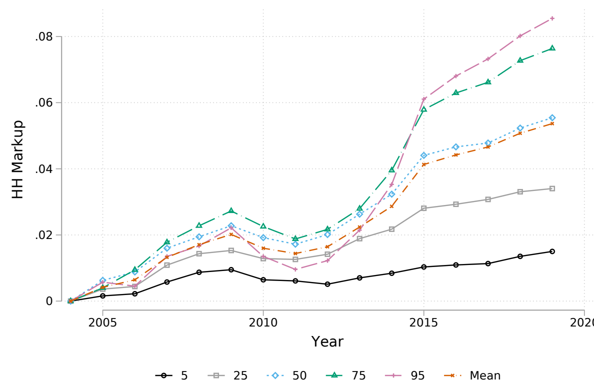
Notes: This figure plots the trends in firm markups from 2004-2019 for the top 10 firms by revenue, both in levels and the magnitude of changes since 2004, assuming Bertrand competition.

Figure G29: Household markups by percentile, 2004-2019

(a) Levels

(b) Changes since 2004



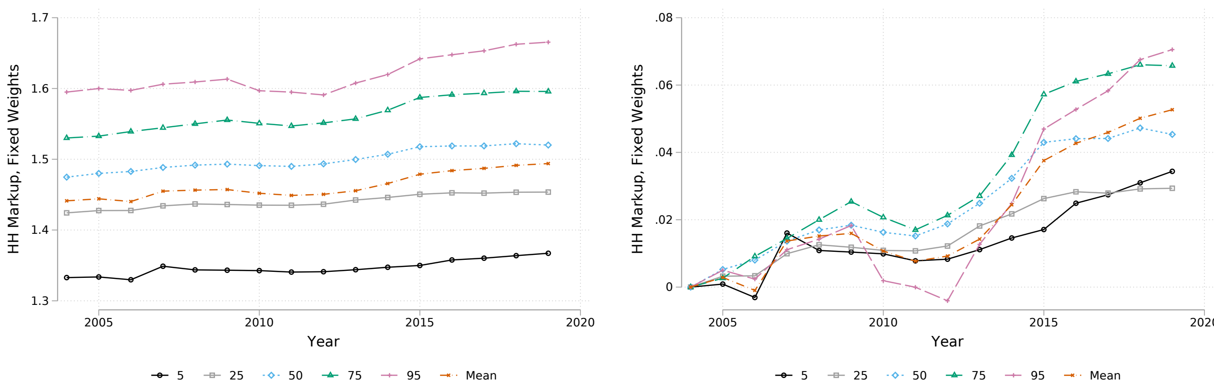


Notes: This figure plots the trends in household markups from 2004-2019 by percentile, both in levels and the magnitude of changes since 2004, assuming Bertrand competition. We calculate markups for each household by multiplying their expenditure share in each firm with that firm's markup, allowing expenditure shares to vary over time.

Figure G30: Household markups by percentile, fixed weights, 2004-2019

(a) Levels

(b) Changes since 2004



Notes: This figure plots the trends in household markups from 2004-2019 by percentile, both in levels and the magnitude of changes since 2004, assuming Bertrand competition. We calculate markups for each household by multiplying their expenditure share in each firm with that firm's markup. We do this both by fixing expenditures at their level calculated across the entire sample period.