The Rise in Household Spending Concentration*

Brent Neiman
University of Chicago

Joseph Vavra
University of Chicago

May 2018

Abstract

(Preliminary and Incomplete)

Household consumption bundles look increasingly different from each other. Using detailed scanner data from 2004-2015, we document that households are concentrating more and more spending on their preferred products. These products, however, are not "superstars" that are purchased by everyone. Rather, household purchases are increasingly idiosyncratic. As a result, aggregate product concentration has actually declined even as product concentration within households has risen. This trend is pervasive across geographic locations and product categories and even holds within demographic and income groups. The growth in household concentration is associated with households purchasing new and dropping old products and is most rapid in retail chains that introduce the most new products. Further, those households with more concentrated product spending pay more for the products they purchase. These patterns suggest firms are increasingly able to introduce customized products or that consumers can better find them, and carry implications for market power and consumer welfare.

*We thank David Argente for providing exceptional research assistance and Tom Wollmann for helpful comments and suggestions. Our analysis is based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from these Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

A growing literature documents increasing fragmentation of culture and digital content (Aguado, Feijoo, and Martinez (2015); Alwin and Tufis (2015)), division by political idiology (Pew Research Center (2014); Gentzkow, Shapiro, and Taddy (2017)), and job polarization and income inequality (Autor, Katz, and Kearney (2006); Piketty, Saez, and Zucman (2016)). This paper documents that, along with these other manifestations of segmentation in modern life, even household retail purchases are increasingly differentiated relative to the national average. We use detailed retail scanner data from 2004-2015 to show that household consumption bundles look increasingly idiosyncratic and to explore implications for concentration, market power, and consumer welfare.

We begin by showing that households are concentrating more and more spending on their preferred products. For example, the average share of spending devoted to a household's top UPC in detailed product categories such as "Carbonated Beverages", "Laundry Supplies", and "Butter and Margarine" has increased by 6 percent from 2004-2015. Increases in household-category Herfindahl indices – a commonly used measure of concentration – are even larger, rising by roughly 10 percent over the same time span. These increases are nearly monotonic across time, highly statistically significant, and robust to a variety of specification changes.

In principle, this increase in household product concentration could be driven by a rising importance of "superstar" brands in "winner take all" industries.[1] In practice, however, we see the opposite. Households purchases are increasingly idiosyncratic. As a result, aggregate product concentration has actually declined even as product concentration within households has risen. Pooling spending for each product category across households, we find that the aggregate Herfindahl has actually *declined* by an average of roughly 20 percent from 2004-2015. We show that the aggregate product Herfind-

[1] Autor, Dorn, Katz, Patterson, and Reenen (2017), for example, argue that this is important for explaining establishment-level trends in sales concentration.

ahl can be formally expressed as the difference between the average household Herfindahl and the cross-household variance of product spending shares, a dissimilarity measure. Rising household concentration is consistent with falling aggregate concentration only if household spending patterns are becoming more heterogeneous, as we find. Together these findings are broadly consistent with forecasts of growing importance of “long-tail" consumption (Anderson (2006)) and product niches and shows that this phenomenon extends beyond e-commerce to broader retail spending.

Our results are not driven by a widening gap between the goods purchased by rich and poor households or between consumers in one region and another.[2] In fact, we find that household consumption bundles are becoming more differentiated even when measuring within geographies, within store chains, and within demographic groups defined by income, race, education, and age. The trend toward more idiosyncratic consumption is pervasive.[3]

So what, mechanically, appears to be driving the shift to more idiosyncratic consumption bundles? The evidence suggests a critical role for new products, either due to the enhanced ability of producers to customize or the expanded search capability of consumers or both. Over our sample period, the typical number of products stocked by retailers has increased significantly.[4] More importantly, the surge in within-household concentration attenuates significantly when we eliminate product churn and apply our measure to a constant panel of goods. That is, the products accounting for rising spending shares for each household within narrow categories are predominantly products which are new to that household. Further, this trend appears most prominently in those stores that have most

[2]This is consistent with the finding in Bertrand and Kamenica (2018) that cultural distance between rich and poor has not grown over time.

[3]We also show that our results are not driven by trends in the share of generic products (Dube, Hitsch, and Rossi (2018)) or by trends in the frequency of shopping and bulk purchasing (Olivier Coibion (2017)).

[4]See Appendix Figure A.1. For example, Yoplait and Dannon have lost substantial market share within yogurts to products from new competitors like Chobani, Fage, and Wallaby. Similar increases in differentiation have also occurred within existing brands: Tostitos chips, for example, now complements its original line of chips with products such as “Scoops” and “Cantina traditional, but most households tend to concentrate purchases in one of these varieties rather than spreading evenly over the three.

significantly expanded their stocked varieties.

Finally, we offer evidence that this enhanced customization of products and concentration of spending carries important implications for pricing, markups, and market power. In particular, this trend toward more idiosyncratic household consumption bundles may imply that a product's aggregate market share is less and less informative about that product's effective market power. To see this, consider two cases. In the first, a product accounts for 5 percent of every household's expenditures. In the second, the product receives 10 percent of the spending of half of the households and none from the rest. The product's seller potentially enjoys greater pricing power in this second case even though the product's aggregate market share is the same as in the first case. To explore such trends more formally, we define product j's "conditional market share" as the total spending on product j divided by the total spending of all households which purchase product j. By construction, conditional market share is at least as large as "unconditional market share" and will be strictly larger for any product not purchased by all households. We find that conditional market shares of products and brands in our data have steadily grown relative to unconditional market shares as cross-household segmentation has increased.[5]

Further, there is a strong positive relationship between household concentration and *relative* prices paid for identical products within a given geographic market. In particular, following Aguiar and Hurst (2007), for each household, we construct an index of the price they pay relative to the economy-wide average price for each product. We then show that there is a strong significant positive relationship between relative price indices and household concentration: households at the 90th percentile of the concentration distribution pay 3-4 percent more for the same products as those at the 10th percentile and households with the largest increases in concentration experience the fastest growth in

[5] To our knowledge, this distinction between conditional and unconditional market shares is first made by Dhyne, Kikkawa, and Magerman (2017) in the context of firm-to-firm trade in intermediate inputs. They provide empirical evidence that conditional market shares are more informative about markups than unconditional market shares.

relative prices. To the extent that marginal costs for a given product are equalized within a market this implies that households with the most concentrated consumption pay the highest markups.

How is it that households with more concentrated consumption pay more for the same goods? One factor we emphasize is that households with more concentrated consumption are relatively less likely to purchase their goods using coupons or during periods with in-store sales, thus suggesting that there are indeed strong interactions between household product concentration, demand elasticities, shopping behavior, and resulting markups.[6] However, it is important to note that this positive relationship between concentration and markups is identified off of cross-household relationships, so it need not imply that absolute markups in the retail sector have increased as household concentration has risen. Our results show that selling to a more concentrated set of households is associated with a relative increase in markups and that concentration has been rising across time, but any common trend in markups across all households will be differenced out of our empirical specification. Thus, it is entirely possible that retail markups have fallen across time despite rising concentration. For example, it is plausible that the introduction of increasingly “niche" products is a way of maintaining profits in an environment with otherwise rising competition and falling markups.

We conclude by noting that our work touches on and draws connections between a number of important themes in recent research. Much like Argente and Lee (2017) and Jaravel (2017), who use detailed micro data to demonstrate that aggregate inflation statistics can miss important heterogeneity in disaggregated data, we demonstrate that aggregate concentration indices may mask interesting changes in spending concentration at the household level.[7] Though we can comment only on concentration and market power in the retail sector, we paint a more complete picture on how these

[6] See also Stroebel and Vavra (2018) for related empirical evidence of such interactions and Berger and Vavra (2018) for evidence of time-varying response of markups to shocks and relationships with market structure.

[7] Our evidence that trends in household concentration are closely related to product innovation complements the argument in Jaravel (2017) that innovation matters for inflation. His analysis focuses on somewhat more aggregated consumption baskets by income groups and a different outcome, so it is largely distinct from our specific analysis. But we reinforce the broader point that product innovation matters for household-firm interactions.

changes manifest and how they might increase profits while also benefiting customers who have greater choice.[8]

2 Data Description and Baseline Sample

We use Homescan data from AC Nielsen to measure household-level shopping behavior.[9] The data set contains a weekly household-level panel for the period 2004-2015. The panel has large coverage, with roughly 160,000 households in over 22,000 zip codes recording prices for almost 700 million unique transactions. The products in the data set cover a large fraction of non-service retail spending. Roughly half of expenditures are in grocery stores, a third of expenditures are in discount/warehouse club stores, and the remaining expenditures are split among smaller categories such as pet stores, liquor stores, and electronics stores.

While panelists are not paid, Nielsen provides incentives such as sweepstakes to elicit accurate reporting and reduce panel attrition. Projection weights are provided to make the sample representative of the overall U.S. population.[10] A broad set of demographic information is collected, including age, education, employment, marital status, and type of residence. Nielsen maintains a purchasing threshold that must be met over a 12-month period in order to eliminate households that report only a small fraction of their expenditures. The annual attrition rate of panelists is roughly 20 percent, and new households are regularly added to the sample to replace exiting households.

Households report detailed information about their shopping trips using a barcode scanning device provided by Nielsen. After a shopping trip, households enter information including the date and store location. They then scan the barcodes of all purchased items. The price is collected in one of two

[8]Barkai (2017) and De Loecker and Eeckhout (2017) calculate stark increases in markups since 1980 using aggregate data, while Traina (2018) and Karabarbounis and Neiman (2018) suggest these increases are far more muted or merely restore aggregate markups toward levels seen in prior decades.

[9]These data are available for academic research through a partnership with the Kilts Center at the University of Chicago, Booth School of Business. See http://research.chicagobooth.edu/nielsen for more details on the data.

[10]We use these projection weights in all reported results, but our results are similar when weighting households equally.

ways: for trips to stores that partner with Nielsen, the average price of the UPC for that store-week is automatically recorded. For trips to stores not partnered with Nielsen, households hand-enter the price paid from their receipt. In addition to the price, households also record whether a product was purchased while “on sale” or using a coupon. In addition, since we know the Universal Product Code (UPC) of each item, information is available on whether a product is generic or name-brand. Products are allocated by Nielsen into three levels of category aggregation: roughly 1304 "product modules", 118 "product groups" and 11 "department codes". For example, "vegetables - peas - frozen" are a typical product module within the "vegetables - frozen" product group within the "frozen foods" department, and "fabric softeners-liquid" is a typical product module within the "laundry supplies" product group within the "non-food grocery" department.

In addition to information on UPCs, Nielsen also provides information on "brands". These are more aggregated than UPCs but still very disaggregated: for example, "Pepsi" and "Caffeine Free - Pepsi" are two different brands, as are "Pepsi" and "Mountain Dew", despite the latter being produced by the same parent company. However, different flavors of Pepsi are typically all listed under the same Pepsi brand. In our baseline analysis, we define a product as a UPC, but we also show robustness results throughout the paper to instead using "brands" as our notion of product. Our baseline defines a product as a UPC code since this is the most fine-grained definition of product available, and will pick up even tiny switches in product purchases across time which may be relevant for changes in concentration arising from the introduction of new flavors or varieties of existing brands. UPCs are also directly assigned by manufacturers and so are less subject to any judgment decisions than the Nielsen defined brand codes which vary somewhat in the level of disaggregation across different product modules. However, there is legitimate concern that UPCs may be too fine a notion of product when considering the concentration of household purchases, since households may view certain

UPCs (for example minor differences in size for otherwise equivalent UPCs) as essentially identical products.[11] For this reason, we show robustness throughout the paper to instead defining a product as a brand rather than a UPC.

Our baseline analysis focuses on annual spending and computes household market shares across products within product groups, but all results are robust to calculating household product market shares in more disaggregated product modules or more aggregated department codes. Since product modules are sometimes added or subtracted from the Nielsen sample, and there is substantial heterogeneity across product modules in the degree of household concentration, our analysis focuses on a set of balanced product modules. This eliminates spurious changes in concentration which might otherwise arise from changes in the set of goods sampled by Nielsen (which do not represent real changes in household's actual consumption and instead merely changes in the categories of consumption reported in Nielsen). This focus on balanced product modules reduces our sample from 118 to 107 product groups. Our analysis excludes fresh produce and other "magnet" items without barcodes since products in these categories cannot be uniquely identified and products with identical product codes in these categories can potentially differ substantially in quality. Our baseline sample includes all households and weights each household using sampling weights provided by Nielsen which are designed to make the Nielsen demographically representative of the broader U.S. population. Appendix Figure A.6 shows that aggregate spending growth in our sample tracks aggregate spending growth in comparable categories in the consumer expenditure survey quite closely. Our conclusions are even stronger when instead using a balanced panel of households to eliminate household composition changes.

[11] It is not clear that we want to classify a switch from spending $10 on Brand-X 64 oz laundry detergent and $10 Brand-X 60 oz laundry detergent to instead spending $20 on Brand-X 64 oz laundry detergent as a large increase in concentration. If UPCs become more homogoneous across time, using UPCs as our notion of product may lead to spurious changes in concentration with no substantive change in household behavior.

Finally, our baseline sample focuses on name-brand UPCs and so excludes generic/private-label products. This is because in order to preserve anonymity of the stores in the Nielsen sample, the exact identity of generic products in the Nielsen data is masked, so all generic UPCs in a product module are typically assigned to one of a small number of anonymized bar-code identifiers. This means that we cannot reliably identify store brands such as Whole Foods-365 15 oz tomato sauce from store brands such as Safeway Signature Select 15 oz tomato sauce. There has been an increase in the private label share of all purchases over the last decade (see e.g. Dube, Hitsch, and Rossi (2018)) so including generic spending which cannot be properly allocated to constituent UPCs might lead to spurious concentration trends. However, we also calculate results including generics and show that this distinction is ultimately unimportant for our quantitative conclusions.

3 Basic Facts and Concentration Decomposition

We begin our analysis by exploring how household product-level concentration has changed across time. For each household i, UPC j, and product group c we calculate total spending $S^t_{i,j,c}$ in year t and associated market share:

$$m^t_{i,j,c} = \left(\frac{S^t_{i,j,c}}{\sum_j S^t_{i,j,c}} \right). \tag{1}$$

Our primary measure of household product concentration is then the Herfindahl of these household-product category market shares:

$$H^t_{i,c} = \sum_j \left(m^t_{i,c,j} \right)^2, \tag{2}$$

It will often be convenient to average these household-category Herfindahls across households and/or categories. To construct a Herfindahl which averages across all household Herfindahls within a category, first define:

$$\alpha_{i,c}^{t} = \frac{\sum_j \left(\omega_i^t S_{i,j,c}^t\right)}{\sum_i \sum_j \left(\omega_i^t S_{i,j,c}^t\right)}, \tag{3}$$

as household i's share of total aggregate spending in category c, where ω_i^t is a household's sampling weight provided to make the Nielsen sample representative of aggregate consumption. We can then define category c's average Herfindahl as the spending weighted average of all household Herfindahls in category c using cross-household weights $\alpha_{i,c}^t$:

$$\overline{H}_c^t = \sum_i \alpha_{i,c}^t H_{i,c}^t. \tag{4}$$

We can then calculate the average household Herfindahl across all categories:

$$\overline{H}^t = \sum_c \alpha_c^t \overline{H}_c^t, \tag{5}$$

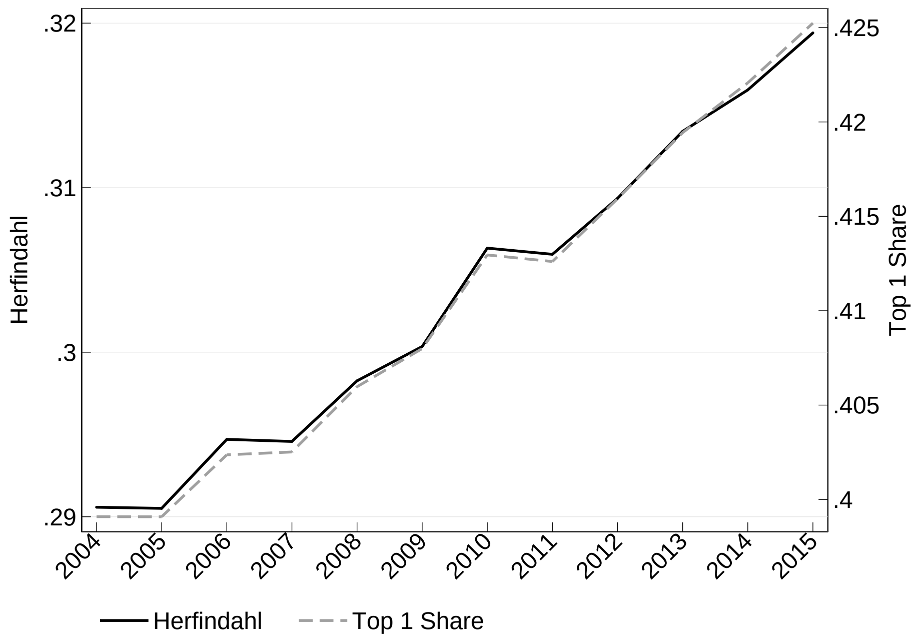
where α_c^t is the share of category c in aggregate consumption.

Importantly, since $\overline{H}_c^t$ varies across categories, $\overline{H}^t$ will change if either $\overline{H}_c^t$ or if α_c^t changes. That is, the average household Herfindahl in the economy can increase either if the household Herfindahl for a particular category increases or if there are compositional shifts in expenditures from low Herfindahl to high Herfindahl categories. Our baseline analysis focuses on within category changes in concentration by fixing $\alpha_c = \sum_t \alpha_c^t$:

$$\overline{H}_{within}^t = \sum_c \alpha_c \overline{H}_c^t, \tag{6}$$

We do so since concentration changes driven by compositional shifts are more mechanical effects that are less likely to be informative about changes in product market power, but we show robustness results which also include these compositional shifts. Figure 1 plots the evolution of $\overline{H}_{within}^t$. In addition to this measure of household-product concentration, we also plot an alternative measure of concentration: the average share of spending by households on their top product in each category.

Figure 1: Household Product (UPC) Concentration



Clearly both measures of concentration increase almost monotonically from 2004 to 2015. The household Herfindahl over this period increases by roughly 10 percent from 0.29 to 0.32. While we do not report standard errors, these means are precisely estimated, and in results below we show that these trends are highly statistically significant. As described in Section 2, these baseline results are computed defining products as UPCs, but Appendix Figure A.3 shows that we see a very similar, though slightly more muted, trend when instead computing concentration over brands rather than over UPCs. Figure 1 holds product group weights fixed across time so that results are not driven by mechanical compositional shifts from less to more concentrated product categories. Appendix Figure A.4 shows that similar patterns obtain when using time-varying weights, although these compositional shifts induce some additional noise since product groups differ substantially in concentration and relative spending on product groups does vary across time.

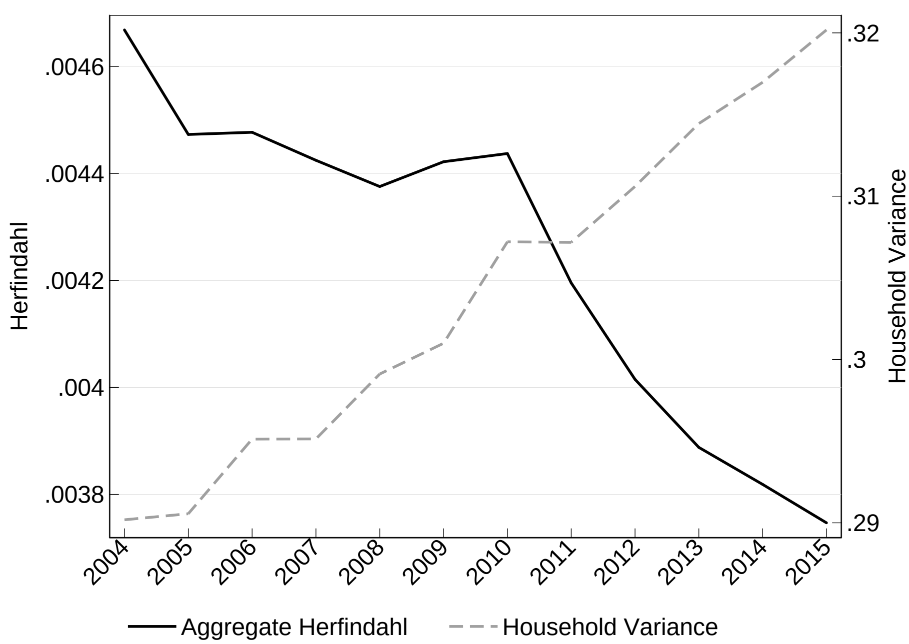
What drives this increase in product concentration across households? A possible explanation would be an increase in the importance of "super-star" products which dominate particular product categories. However, we now show our second fact, that at the same time household-product concentration has been rising, there has been a striking decline in aggregate product-concentration. In particular, define the aggregate market share of product j in a category c as:

$$m^t_{j,c} = \frac{\sum_i \left(\omega^t_i S^t_{i,j,c}\right)}{\sum_i \sum_j \left(\omega^t_i S^t_{i,j,c}\right)}, \tag{7}$$

and the aggregate Herfindahl in category c as:

$$H^t_{agg,c} = \sum_j \left(m^t_{j,c}\right)^2, \tag{8}$$

Figure 2: Aggregate UPC Concentration and Cross-Household Variance



Just as with the household Herfindahl, we can average this across categories weighting by category spending. Figure 2 shows that this average aggregate Herfindahl has declined substantially over the last decade, falling by almost 20 percent. This pattern is fundamentally at odds with a super-star product model of market concentration: at the aggregate level, consumption is becoming less concentrated over time, not more. Indeed, we can formally relate fact 1 and fact 2 with the decomposition (Radaelli and Zenga, 2002):

$$H^t_{agg,c} = \overline{H}^t_c - \overline{V}^t_c, \tag{9}$$

where

$$\overline{V^t_c} = \sum_i \alpha^t_{i,c} \left(\sum_j \left(m^t_{i,j,c} - m^t_{j,c} \right)^2 \right) \tag{10}$$

measures cross-household variance in household-product market shares. Intuitively, if all households consume identical product bundles then $\overline{V}^t_c = 0$ and household concentration and aggregate concentration are identical. Holding household concentration constant, an increase in cross-household variance leads to a decline in aggregate concentration.

On average, the cross-household variance $\overline{V}^t_c$ is large so that the aggregate Herfindahl is much lower than the average household Herfindahl. Put differently, within categories households typically only consume a few products while aggregate category spending is instead typically split over hundreds of products. Figure 2 shows that household variance has increased substantially across time. Indeed, this cross-household variance term must increase by more than the increase in household concentration in order to explain the decline in aggregate product concentration. This means that even though households are increasingly concentrating their individual consumption, they are not concentrating on the *same* set of products: polarization and fragmentation in product-level consumption is increasing across time. The increase in household concentration is not a story of superstar products.[12]

[12]Our results are distinct from studies of trends in production concentration using Census data. First, our spending categories are predominately limited to groceries. Second, we measure concentration over products and brands rather than firms or establishments, which typically produce many products and brands. In principle we could link products to producers using barcode information and could then compare production concentration in our data to the flat path of wholesaler concentration in Census data computed by Autor, Dorn, Katz, Patterson, and Reenen (2017). However, this cannot be done in practice because store-brand products (“generics") are anonymized in our data and so cannot be linked to producers. While we can make various assumptions on the allocation of products from firms (from ignoring them to

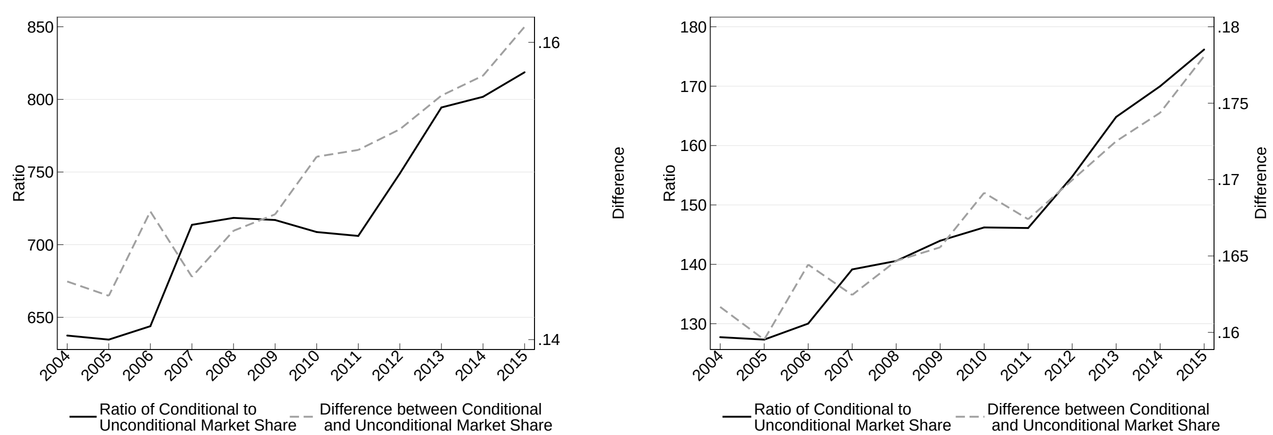
We next offer a closely related but slightly different perspective on these increasing household concentration and segmentation trends. In particular, we explore how the distribution of total demand for each product is distributed across different households by computing the relationship between what we call “unconditional" and “conditional" market shares. The unconditional market share $m^t_{j,c}$ is simply product j's share of *all* spending in category c, as defined in equation (7) above. By contrast, we define the conditional market share $c^t_{j,c}$ as product j's share of the *subset* of spending in category c done by households which purchase product j:

$$c^t_{j,c} = \frac{\sum_{i \in \Omega} \left(\omega^t_i S^t_{i,j,c} \right)}{\sum_{i \in \Omega} \sum_j \left(\omega^t_i S^t_{i,j,c} \right)}, \tag{11}$$

where Ω is the set of households with non-zero spending on product j. Note that a product's conditional market share only differs from its unconditional market share because the denominator excludes spending by households that do not buy it. Conditional market shares, therefore, are weakly larger than unconditional market shares.

Figure 3: Conditional Vs. Unconditional Product Market Shares

(a): UPC **(b):** Brand



The left panel of Figure 3 compares the conditional and unconditional market shares of products

treating them as produced by a single firm), this proves critical for conclusions on production concentration (though not for conclusions on household spending concentration which is the focus of our analysis). We can measure concentration of spending at retailers and find it has increased as in Autor, Dorn, Katz, Patterson, and Reenen (2017).

over time. The solid line of that panel plots a weighted average across products of their ratio and demonstrates that the conditional market share has increased proportionately more than the unconditional market share, while the dashed line shows that the average across products of their difference has also steadily increased since the beginning of the sample. The right panel of Figure 3 shows these same patterns hold if we instead measure them as brands.[13] To the extent the households that consume a product constitute its relevant market, rather than all consumers in the economy, these plots demonstrate that products' shares of their market have increased significantly relative to what one would infer from aggregate data alone. Indeed, such a divergence is required to reconcile our findings of declining aggregate concentration and rising concentrating in household spending bundles. To the extent that effective market power depends not just on total demand but also on its distribution across households, this may have implications for changing market power and resulting markups, as we explore below.[14]

4 Zooming In: Particular Products and Markets

We have documented that household concentration and segmentation are increasing, and this has caused an increase in conditional relative to unconditional market shares. We now show that this phenomenon is pervasive throughout the economy and is not driven by particular products or product groups.[15]

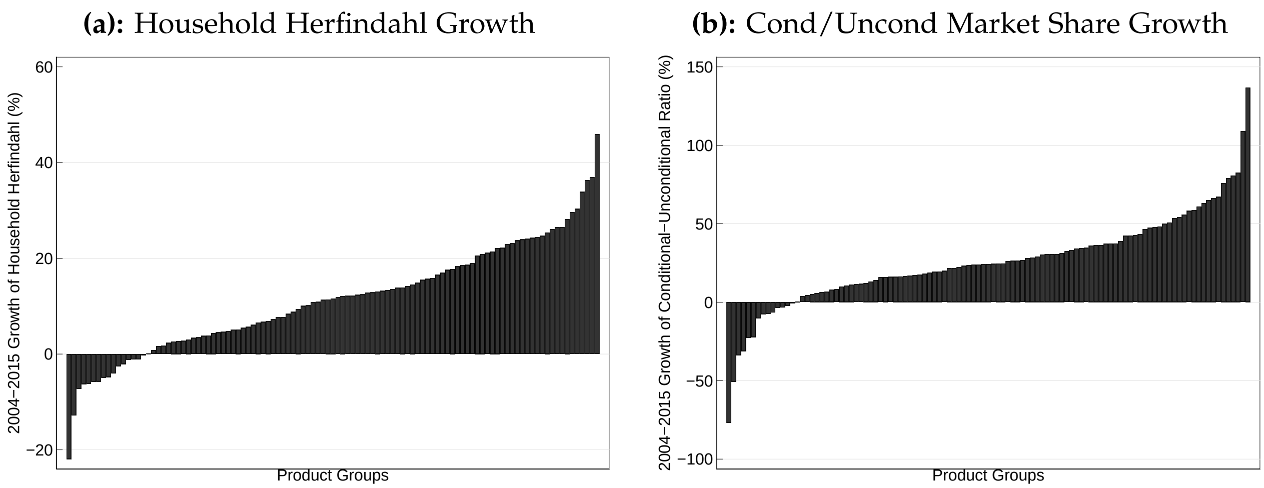
Figure 4 shows that the increase in household concentration and the increase in conditional rel-

[13]We weight the ratios with the unconditional market share since otherwise the averages are dominated by items with tiny aggregate market shares. We do not weight the differences since observations with large values of $m^t_{j,c}$ already naturally receive more weight in this calculation.

[14]See Dhyne, Kikkawa, and Magerman (2017) for an analysis of intermediate input trade among Belgian firms, which is to our knowledge the first study of such conditional market shares. They demonstrate large differences between the conditional and unconditional measures and provide evidence that, of the two, a firm's conditional market share is more closely related to its inputed average markup.

[15]Our baseline results exclude generic products since generic UPCs are not typically uniquely identified in Nielsen data. However, Appendix Figure A.2 shows that the generic vs. name-brand product distinction is not important for our results by redoing our calculations including generic UPCs.

Figure 4: 2004-2015 Growth For Each Product Group



ative to unconditional market shares is broad-based and appears in nearly every product category. The mean and median product groups exhibit cumulative growth in their household Herfindahls of around 11 percent between 2004 and 2015, while the mean and median growth of the ratio of conditional to unconditional market shares is 26 percent and 24 percent. 91 out of 107 product categories exhibit increasing household concentration and 93 categories exhibit increasing ratios of conditional to unconditional market shares. Thus, the aggregate trends are not driven by any particular category and are instead broad-based.

There also are not any particularly strong common patterns differentiating categories with high and low conditional-unconditional ratio growth. The three categories with the lowest growth of conditional-unconditional ratios are greeting cards, photographic supplies, and ice. The three categories with the highest growth rates are disposable diapers, detergents, and cereal. The three categories with lowest household concentration growth are coffee, snacks-spreads-dips, and vitamins. The three categories with highest growth are stationary, wrapping materials, and ethnic health and beauty products. Overall household concentration growth is negatively correlated with spending growth in a category with an elasticity of -0.13 and t-stat of -6.85. This is not particularly surprising since most products can only be purchased in whole units. This means that a category in which a household rarely purchases is likely to exhibit high concentration. Conversely, the elasticity of the conditional-unconditional ratio to category spending growth is 0.08, although this relationship is insignificant. However, we now provide some evidence that these relatively mechanical spending growth effects do not drive our overall trends and then show this in our household level regressions more explicitly below.

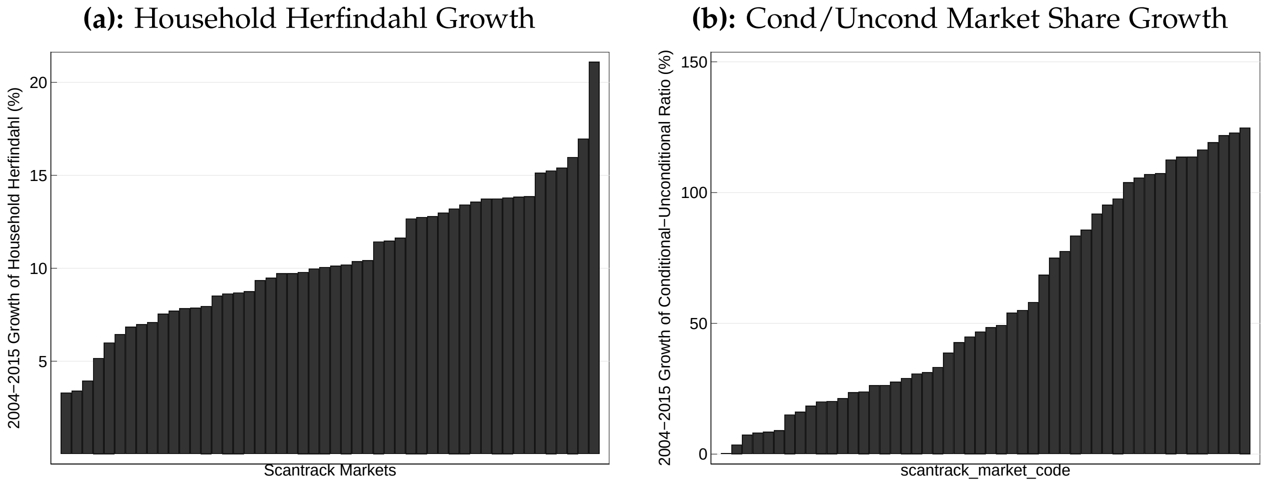
To further demonstrate that spending growth is not the primary determinant of the concentration patterns we document, we run the following category-level regression:

$$\log(\overline{H}_c^t) = \beta t + \alpha \log(s_{c,t}) + \delta_c + \epsilon_{c,t}, \tag{12}$$

where $\overline{H}_c^t$ defined above is the average household Herfindahl in category c in year t, $s_{c,t}$ is the average spending in this group-year and δ_c is a product group fixed effect. The coefficient β, then gives the average annual growth rate of concentration after controlling for mechanical spending effects captured in the coefficient α. Running this regression yields $\beta = .0103$ (t-stat 11.25) and $\alpha = -.13$ (t-stat -4.2) while running the same regression with no controls for category spending yields $\beta = .0097$ (t-stat 11.25). Thus, not only are our results not driven by mechanical spending effects, controlling for category spending changes actually mildly strengthens the overall trend. When we turn to household level controls in the following section we will show that similar results also obtain when controlling for household level spending rather than category averages.

Thus far, all results have shown national trends and ignored geographic variation. Do these same results hold within different cities or could they be driven by some cross-city compositional effect? Figure 5 recalculates the growth rate of household Herfindahls as well as the growth rate of conditional-unconditional market share ratios for each scantrack market in the Nielsen data set.

Figure 5: 2004-2015 Growth For Each Scantrack Market

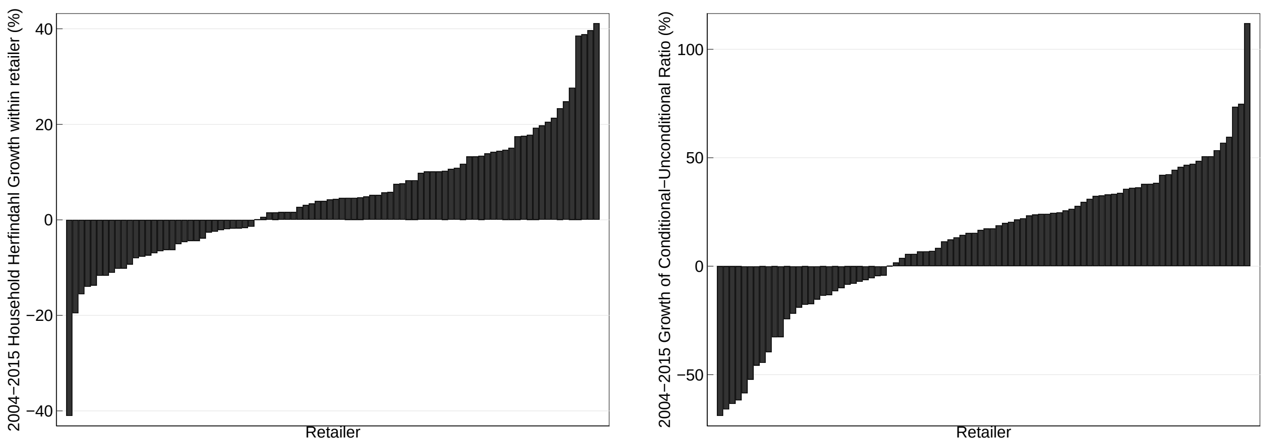


The growth rates of both household concentration and the ratio of conditional to unconditional market shares are positive in every single scantrack market, so the upward trend is not simply driven by compositional changes. While there is variation across cities, it does not appear to be correlated with broad regions or other obvious city characteristics. The four cities with the highest growth of household concentration are Albany, Milwaukee, Louisville, and Detroit while the four lowest are Little Rock, Raleigh, Memphis, and Nashville. The four cities with the highest growth rates of conditional-unconditional market shares are Omaha, Grand Rapid, Richmond, and San Diego while the four lowest are Columbus, Buffalo, St. Louis, and Charlotte. Overall, there is some evidence of convergence in both household concentration and conditional-unconditional ratios as the cities with the lowest initial values tend to grow more quickly. The cross-city coefficient of variation of concentration has declined 28 percent and the cross-city coefficient of variation of conditional-unconditional ratios has declined by 35 percent. This means that across the US, cities are becoming more similar in their concentration patterns: all cities have growing household concentration, but those with initially lower concentration levels are seeing concentration grow more quickly.

One might also wonder whether these patterns are driven by cross-retailer differences. Perhaps households are increasingly shopping in one particular retailer and this is driving the increase in household concentration if the set of products available in a single retailer is lower than that available across multiple retailers. To explore this, we recompute household product concentration but measuring market shares within a category *and* retailer pair instead of aggregating a household's spending across all retailers within a category. That is, we calculate within-retailer concentration instead of concentration across purchases made at all retailers.[16] Figure 6 shows that while there is a little more heterogeneity, household concentration is rising within most retailers, so the upward trend is not primarily about where households are shopping but is instead about changes in the composition of their purchases within retailers. We return to an analysis of this heterogeneity below.

Figure 6: 2004-2015 Growth Within Retailers

(a): Household Herfindahl Growth **(b):** Cond/Uncond Market Share Growth



5 Concentration Trends and Demographic Variables

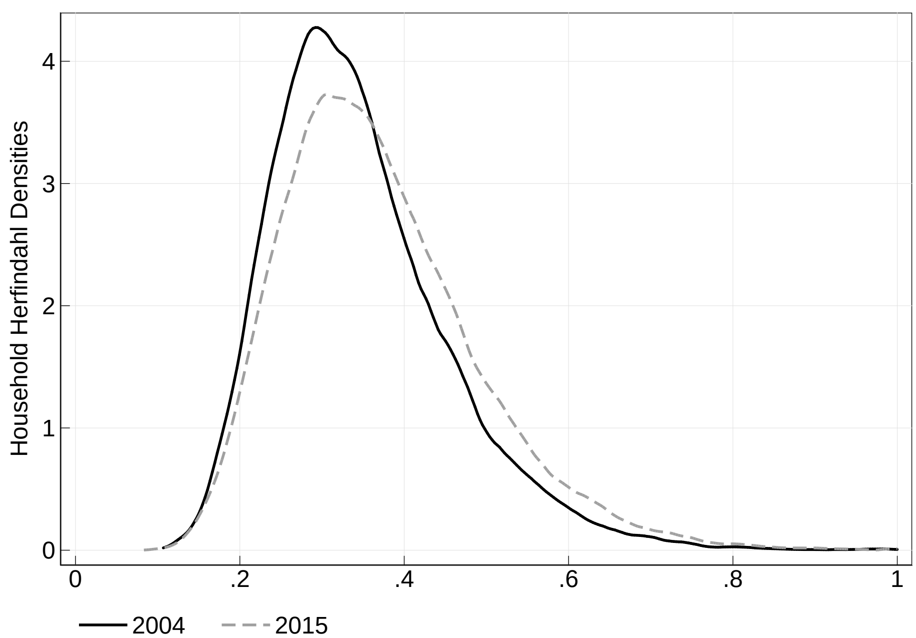
For now we have focused only on trends in the data averaging across many households. We now turn to a more disaggregated analysis of household concentration to make several points: 1) Household

[16]Nielsen does not track a constant set of retailers across time and retailers are sometimes added and subtracted for institutional reasons. To eliminate spurious effects due to sample entry/exit, we restrict to a balanced panel of retailers. We also restrict to retailers in which at least 500 households shop annually in order to ensure accurate measures of within-retailer concentration measures. Results are similar with higher and lower thresholds. While this baseline restriction drops around 90 percent of retailers, it only eliminates around 30 percent of spending.

concentration differs dramatically across households and the level of concentration is correlated with a variety of demographic variables. 2) Controlling for these demographic variables has almost no effect on the aggregate trend, so that this trend is not explained by changing demographics across time. 3) Increasing household concentration holds robustly within a broad-swath of demographic groups.

Figure 7 shows the large amount of heterogeneity in concentration across households using a kernel density plot of average household Herfindahls. For each household i and year t we compute its average (expenditure weighted) Herfindahl overall all product categories $\overline{H}_i^t$, and the figure shows the density of $\overline{H}_i^t$ in 2004 and 2015.[17] Two points are apparent: 1) There is a huge amount of heterogeneity across households. 2) The distribution shifts to the right between 2004 and 2015.

Figure 7: Variation in Concentration Across households



What explains the large differences in concentration across households? To explore this, we regress

[17]Note that $\overline{H}_i^t$ is analogous to $\overline{H}_c^t$ defined in equation 4 but now averaging household-category Herfindahls for each household i over all categories instead of averaging overall all households for each each category c.

household Herfindahl levels on a variety of observables.[18] To control for compositional effects to make results most comparable to our initial figures, we run this regression at the *household* $\times$ *productgroup* level with product group fixed effects, and results are clustered by household and product group. Column 1 of Table 1 shows that there is a strong negative relationship between household concentration and household size. This is not surprising since larger households have more internal heterogeneity and are purchasing products for more individual people so they split consumption in categories across a broader mix of products. Since this effect is quite strong and relatively mechanical, all remaining specifications also control for household size. While the R^2 on this regression is 0.32, it is important to note that this is mostly explained by the product group fixed effects since a regression of household-product group-year Herfindahls on product group fixed effects alone gives an R^2 of 0.3087. Thus, the usual caveat about interpreting incremental R^2 from any particular demographic control in the presence of fixed effects applies. In column 2 we show that there is an age profile of concentration, with Herfindahls growing steadily over the life-cycle. Column 3 introduces income controls, but and shows this has little effect on concentration.[19]

In column 4 we additionally introduce controls for household spending. Similar to the cross-category results, higher spending is strongly associated with lower concentration. Again, this is not particularly surprising since individual varieties must be purchased in discrete units, and similar patterns would also arise under preferences with some love for variety. Column 4 also shows that after controlling for spending, the age profile of concentration is much stronger. Similar to the effects of household size, effects of spending on concentration are relatively mechanical and so we control for household spending in most of our remaining specifications. While spending is clearly endogenous, it

[18]These results in levels are very similar to those in Hansen and Singh (2015), but our focus is largely on trends.

[19]Appendix Table A.1 which repeats these regressions at the household-level with no product group fixed effects shows some modest effects of income. This implies that effects of income on household concentration are largely driven by category composition effects rather than within category concentration changes.

is important to note here that we are not attempting to give any particular causal interpretation to these relationships and are instead merely trying to statistically decompose the overall variance of household Herfindahls and assess the extent to which the trend documented in the previous section can be explained by demographics or other observables. In Column 5, we introduce a variety of additional controls available in the Nielsen data including location, race, education and type of residence.[20] While some of these coefficients are significant, there are no particularly interesting patterns, so we suppress these particular coefficients. Comparing the R^2 in columns (4) and (5) shows that these additional controls have little additional explanatory power.[21]

To what extent do these same controls matter for the trend increase in household concentration documented in the previous section? Can changing demographics explain this trend? For example, given the age profile shown in Table 1, can things like the aging population in the US potentially explain our trend? Table 2 shows that these observable controls have little effect on the concentration trend, so it is not explained by changes in these characteristics across time. To show this, we simply re-run the regressions in Table 1, but with an additional linear time trend. The key takeaway is that the trend is large, highly significant and largely insensitive to including these observable controls.[22] With the full set of controls in column (6), the coefficient on the linear time-trend is .0035.[23] Using this coefficient implies that from 2004-2015 even after controlling for a variety of observable demographics, average household Herfindahls have grown by 11*.0035=.0385. This is roughly a 13 percent increase from the initial level in 2004.

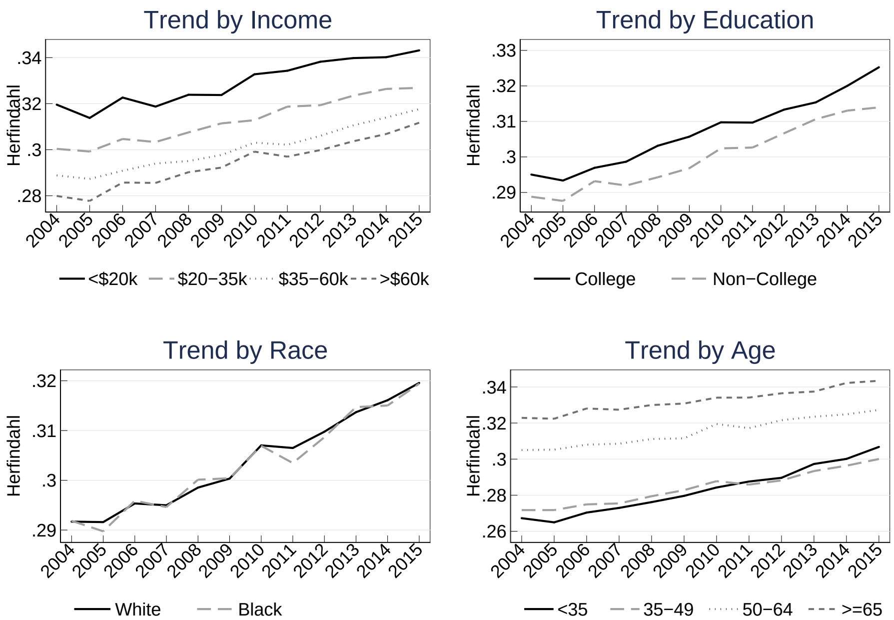
[20]Including a "kitchen-sink" of all controls available in the Nielsen data offers little additional predictive power but substantially increases the computational burden in the regressions.

[21]Since demographic variables are often highly correlated, comparisons of the explanatory power of different variables should be made with caution. For example, the incremental R^2 from household size is diminished after including other demographic controls.

[22]We have also run specifications where we control for the fraction of spending on bulk items or for the number of shopping trips made by households, as in Olivier Coibion (2017), and this has no effect on our results. This means the trends they document are unrelated to our spending patterns.

[23]Note, to improve readability and limit leading zeros, the table lists the coefficient from a regression on $t/100$ rather than t.

Figure 8: Trends by demographic group



These results show that simple compositional changes in sample demographics across time are not explaining the trend increase. But do these same trends hold within each demographic group, or could they be driven by some groups with particularly strong trends? To look at this, we look at trends within different groups. We start with the simplest analysis, by simply recomputing the Herfindahl trend in Figure 1 for a variety of different demographic groups. That is, we just split the sample by a single demographic variable at a time and recompute trends. Figure 8 shows there are clear level differences which mirror those shown in the previous Table 1, but the the upward trend is similar across each of these splits.

The results in Figure 8 do not account for the fact that many observable factors vary simultaneously across demographic groups. However, Table 3 shows that calculating within-group trends after allowing for a variety of additional demographic controls in levels does not change the conclusions.[24]

[24]For comparability with earlier tables, we split by finer demographic groups than in Figure 8.

In particular, we estimate $\overline{H}_i^t = \alpha + \gamma_d t \times D_i + D_i + \xi_{i,t} + \epsilon_i, t$, where D_i is an indicator variable that i is in a particular demographic group of interest, and $\xi_{i,t}$ is an additional vector of controls for household i. The coefficient of interest is γ_d, which measures the Herfindahl time-trend for demographic group D after controlling for a variety of other demographic characteristics in levels.[25] Table 3 shows that within-group trends are similar and significantly positive even after controlling for other observable household characteristics, so the basic patterns in Figure 8 continue to hold.

6 The Role of Relative Prices

Why do we care about the rise of household concentration and segmentation? It is potentially another relevant dimension of "effective" market power above and beyond more traditional measures of aggregate within-sector concentration, and our results show that over the last decade at the product-level these statistics have moved in opposite directions. Are these concentration trends then related to pricing power and household inflation? One might expect that pricing power could be higher for a product with 100 percent market share in a category for 1 percent of households than for a product with 1 percent market share for 100 percent of households, even though these two products would have identical aggregate market shares. To the extent that variety differentiation is leading to increased sorting and more captive households, one might expect this to be a force towards greater pricing power and higher prices. However, it is also equally possible that as the product space becomes more finely differentiated that the elasticity of demand facing any one product actually rises. Ultimately, it depends on whether households view the increase in variety as making a wider set of products more substitutable or whether households increasingly prefer their purchased product over all others. While estimating a formal demand system for every household and product is beyond the

[25] All regressions include the full set of controls used in Tables 1 and 2: dummies for household size, age groups, income groups, scantrack markets, type of residence, race, as well as the log level of spending. To make the interpretation of γ_d easier, we do not include a separate time-trend not interacted with t; that is we have no excluded group in the trend term. This means that γ_d can be interpreted as each group's trend without adding in another common cross-group term.

scope of this paper, we can look at the relationship between relative prices paid and household concentration to get some sense of whether the rise in household concentration is associated with changes in effective pricing power.

To do this, we first construct a household relative price index $\hat{P}_i^t$ for each household i in year t as in Aguiar and Hurst (2007). In particular, for each UPC j and year t, we compute the average expenditure weighted price $\overline{p}_j^t$ across all households in each geographic scantrack market.[26] We then compute the deviation between $\overline{p}_j^t$ (computed over all households) and the average price for product j paid just by household i: $\hat{p}_{i,j}^t = p_{i,j}^t / \overline{p}_j^t$. Finally, we compute a relative price index for each household i by taking a household expenditure weighted average of individual product relative prices: $\hat{P}_i^t = \sum_j \frac{S_{i,j}}{\sum_j S_{i,j}} \hat{p}_{i,j}^t$.

We then run the regression $log\hat{P}_i^t = \beta \overline{H}_i^t + \xi_{i,t} + \delta_i + \epsilon_i, t$ where $\xi_{i,t}$ is a vector of potential controls identical to those in the tables above exploring trends, and δ_i is a household fixed effect which we can either include or exclude from the regression. When these fixed effects are included, the regression is identifying the relationship between changes in household concentration and changes in household relative prices across time within a household. With no fixed effects, the regression is instead primarily identified off of level differences across households. Table 4 shows the semi-elasticity β and implies that there is a strong positive relationship between household concentration and household relative prices. This holds both with and without household fixed effects, and is stronger after controlling for a variety of additional demographic variables. In the specifications with a full set of additional controls, the semi-elasticity ranges from 0.103 to 0.162. The 10th percentile of $\overline{H}_i^t$ is 0.21 and the 90th percentile is 0.47, so these estimates imply that households at the 90th percentile of concentration pay roughly 2.7-4.2 percent more for the same products.

Note that this relative price index computes the prices paid for particular UPCs relative to the

[26] Since each household is located in a single scantrack market in each year, we simplify notation by not separately indexing prices by i and s.

average price paid for those same UPCs in the economy, so this positive relationship between relative prices and concentration cannot be driven by compositional differences in which households purchase which products. Instead, the interpretation of these relative price effects is that a household which has more concentrated consumption will spend more for the same UPC than a household with less concentrated consumption. How is that households with more concentrated consumption pay more for identical products than households with less concentrated consumption? Given the nature of the packaged good retailers in our data, it is unlikely that this is arising through explicit price discrimination at a point in time across households. This type of explicit price discrimination is especially unlikely since our results hold even when comparing household concentration and relative prices within a given retail chain. Instead it is more likely to be driven by intertemporal price discrimination through sales, coupon use and differences in shopping effort.

To provide some evidence that this is the case, the first four columns of Table 5 show that greater household concentration is associated with a decline in the fraction of purchases made using coupons and on items which are on sale. While interesting, this fact alone could occur because more concentrated households are more likely to purchase the same UPC using coupons than less concentrated households, or because more concentrated households tend to purchase products which just issue more coupons and so are generally purchased using coupons more often. That is, this relationship could be driven by (1) differences across households in coupon use for a given product or by (2) differences across households in the set of products purchased. Columns (5)-(8) show that effect (1) is important by showing there is a strong *within* product decline in coupon and sales intensity with concentration. We measure this by first calculating the top UPC j by market share purchased by each household i in each product group and then calculating the fraction of purchases of product j by household i which are made using coupons and on sale. We call these respectively $frac^{c}_{i,j}$ and $frac^{s}_{i,j}$.

We then calculate the average fraction over all households of purchases of product j which are made on sale or using coupons, which we call $\overline{frac_j^c}$ and $\overline{frac_j^s}$. Finally, for each household i we then calculate $diff_{i,j}^c = frac_{i,j}^c - \overline{frac_j^c}$, average this over all product groups for the household and then regress this on concentration. In essence, this measures whether a given household is more or less likely to purchase its top product using coupons or on sale than is typical for that same product more generally. Table 5 shows that there is a strong negative relationship between these sales/coupon intensity measures and concentration, which means that more concentrated households are less likely to purchase the exact same product on sale or using coupons than the average household.

How important is this negative relationship between concentration and coupon/sales use for the concentration-relative price relationship documented in Table 4? To assess this, we again regress relative price indices on household concentration but now include additional controls for the average fraction of products each household purchases on sale and using coupons. Table 6 shows that coupon and sales use explain roughly half of the concentration-price relationship.[27] This is a relatively large effect given the fairly simple nature of these coupon/sales measures which make no adjustment for the size of coupons or sales as measures of intensity. We are unable to measure the effect of shopping composition across stores within geographic areas since there are too few household $\times$ store observations within each market to accurately calculate store-level price indices, but we suspect that some fraction of the remaining price-concentration relationship is driven by differences in shopping behavior across stores.

[27] Note that column (1) in Table 6 is identical to column (1) in Table 4 and column (3) in Table 6 is identical to column (2) in Table 4.

7 The Role of Variety and Product Innovation

What drives these results? Concentration can change by households 1) changing the set of products they purchase (extensive margin effects), 2) changing the prices they pay for products which they purchase continuously across time (intensive margin price effects), or 3) changing the quantities they purchase for products which they purchase continuously across time (intensive margin quantity effects). We now show that changes along the extensive margin are particularly important for these trends. For these results, we move to a panel specification since we require at least two time periods for a household to measure which products for that household are new and which are continuing.[28] In particular, for each household i which is observed in t and $t+1$, we call a product "continuing" if it is purchased by that household in both t and $t+1$.[29] We can then calculate the Herfindahl for household i over all products and separately over the subset of continuing products in t and $t+1$. To calculate trends, we calculate the differences between t and $t+1$, accumulate these across time and pin down the level using the initial period.[30]

Figure 9 shows the upward trend is much stronger when using all UPCs than when restricting to continuing products, implying a large role for product entry and exit in generating concentration increases.[31] While we can measure UPC entry and exit precisely in Nielsen data, one might be concerned that measuring product turnover at the UPC-level may overstate the degree of product entry

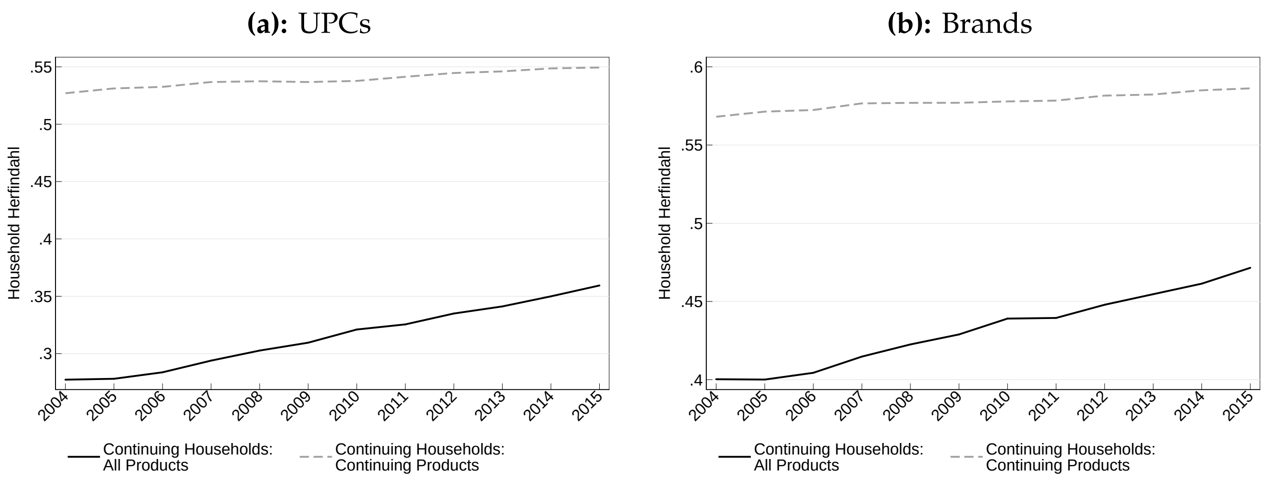
[28] Alternatively, we can define "new" and continuing products in the aggregate, and then recompute our baseline specification in Figure 1 separately for new and continuing products in the aggregate. This does not require identification off of the household panel element, but it also does not fully isolate extensive and intensive margin effects at the household level since a product may be new to the household between t and $t-1$ but not new to the aggregate. Nevertheless, Appendix Figure A.5 shows that the results are fairly similar, finding a large role for new products in the aggregate driving concentration trends.

[29] When defining products as UPCs, continuing products make up around 30 percent of transactions and 40 percent of spending. When defining products as brands, continuing products make up 40 percent of transactions and 50 percent of spending.

[30] Note that only results in changes and not in levels are well-defined, since this specification produces two values of the Herfindahl for each year, one corresponding to its treatment as a base year and another corresponding to its treatment as the continuation year.

[31] Note that the trend for "all-products" is larger than in Figure 1, since this analysis is using within-household variation. See Appendix B for discussion.

Figure 9: 2004-2015 Household Herfindahl growth for continuing products



and exit relevant to households. In particular, UPCs will often change when there are trivial changes in packaging or other characteristics. This could result in us calling a product new even if a household might instead perceive it as a continuing product. This is less of a concern for brands, so the right panel repeats this calculation defining a product as a brand rather than UPC. Results are similar, with the upward trend again substantially attenuated when restricting to continuing brands.

These results imply that increases in concentration along the “intensive" margin for continuing products is relatively modest and that most of the action happens on the extensive margin. Nevertheless, despite the small overall intensive margin effect, it is still interesting and feasible to further decompose it. In particular, for continuing products we can explore the relative importance of intensive margin effects arising due to changing prices vs. changing quantities.[32] To explore this, we compute two alternative versions of the “continuing" product Herfindahl line in Figure 9 In the “constant price" version, we recompute the Herfindahl change for each household on the set of continuing products between period t and period $t+1$ but instead of using period $t+1$ prices to compute the Herfindahl in $t+1$, we instead using period t prices. Similarly, the constant quantity version holds

[32]Note that we cannot explore this decomposition for non-continuing products since we by construction have no pre-period comparison that we can split into P vs Q changes.

quantities constant at base year values when computing Herfindahls in $t+1$. We focus only on UPCs since the price for a "brand" is not well-defined as a given brand can encompass substantial changes in UPC-price composition across time.

Figure 10: Intensive Margin P v. Q effects for UPCs

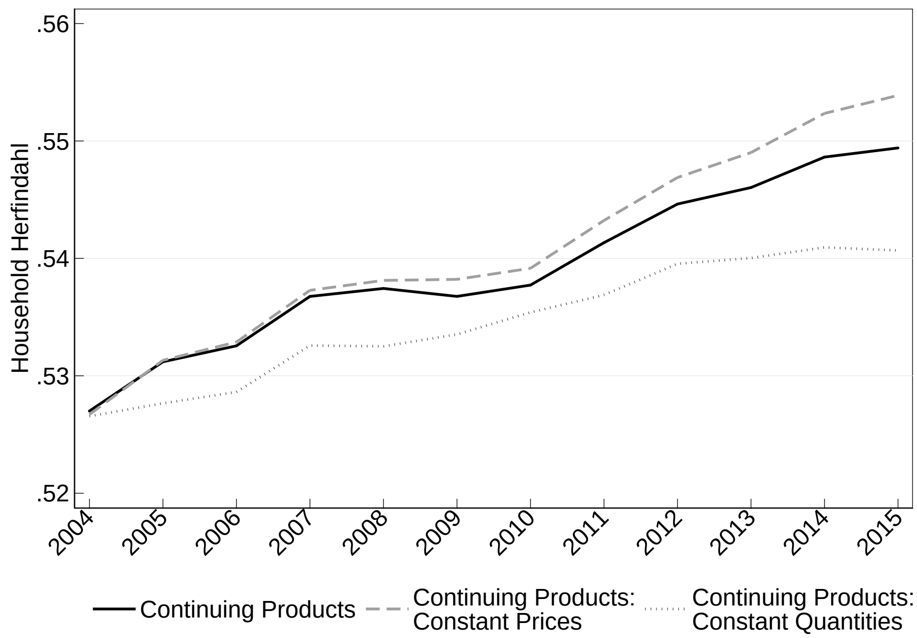


Figure 10 shows that changes in the quantity of demand are more important for driving changes in concentration across time than are changes in prices. That is, increases in market share are predominantly driven by the quantity of demand for the top products rising, not by paying a higher price for the same quantities of those products. The fact that the constant price trend is stronger than the continuing products trend with both price and quantity variation means that there is a negative covariance trend in that products with higher quantity growth within the household are experiencing lower price growth. This might seem surprising in light of the results in the previous section, but these results are not inconsistent, and it is not surprising that there is a negative covariance term between P and Q since this would be arise in the presence of standard substitution effects.

To reconcile the results, first recall that the previous section focused on the relative price for all products purchased by a household while these results are only measuring prices for continuing products. More importantly, in this decomposition between P and Q, what matters is the relative price across products *within* a given household across time, while in the previous results, we are measuring the relative price *across* households. An increase in the price of the overall consumption basket for high concentration households relative to low concentration households need not imply that the relative price for the highest market share item for the high concentration household rises relative to its lowest market share item.[33]

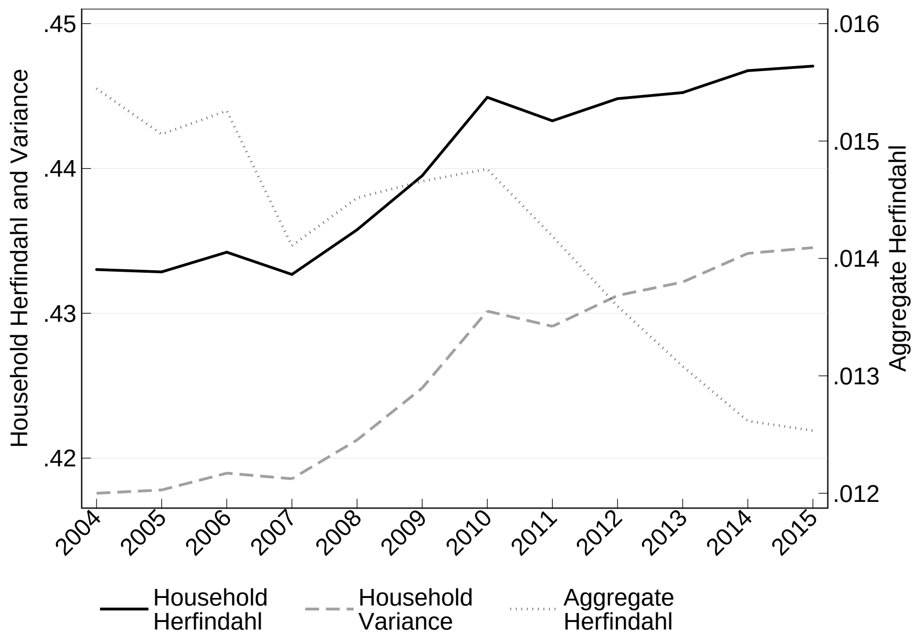
Returning to the overall trend, the above results suggest that changes in varieties along the extensive margin are important for understanding the trend increase in household concentration and that results are not driven simply by increasing taste heterogeneity amongst a constant set of products. We next differentiate two margins along which variety segmentation might occur. First, retailers may be increasingly specializing by stocking different products from each other. If households consistently shop at different retailers then this could explain increasing household concentration and segmentation. Second, retailers may instead be expanding product availability and households shopping at a given retailer may be sorting themselves into particular narrow products which best fit their preferences.

Under the first explanation, conditional on the choice to shop at a particular retailer, households are facing more narrow choices. We would then expect that the increasing concentration would hold within retailers but that cross-household variance within retailers would be declining. Under the second explanation, household concentration and variance should both rise even within retailers.

[33]For example, if household A and B both initially consume 1 unit of product 1 and 2 at price \$1, and then household A consumes 2 units of product 1 at price \$2 and one unit of product 2 at price \$3, household A's concentration rises, the price of A's consumption basket relative to B rises, but the relative price of A's top product 1 declines relative to product 2 so that when decomposing into P and Q effects, the P effect is negative. This example thus delivers exactly the qualitative results found in these two sections and so shows they are conceptually consistent.

While we have already shown in Figure 6 that household concentration rises within retailers, this could be consistent with either mechanism. Thus, we now recompute cross-household variance as well as “aggregate" within-retailer Herfindahls, as in Figure 2.

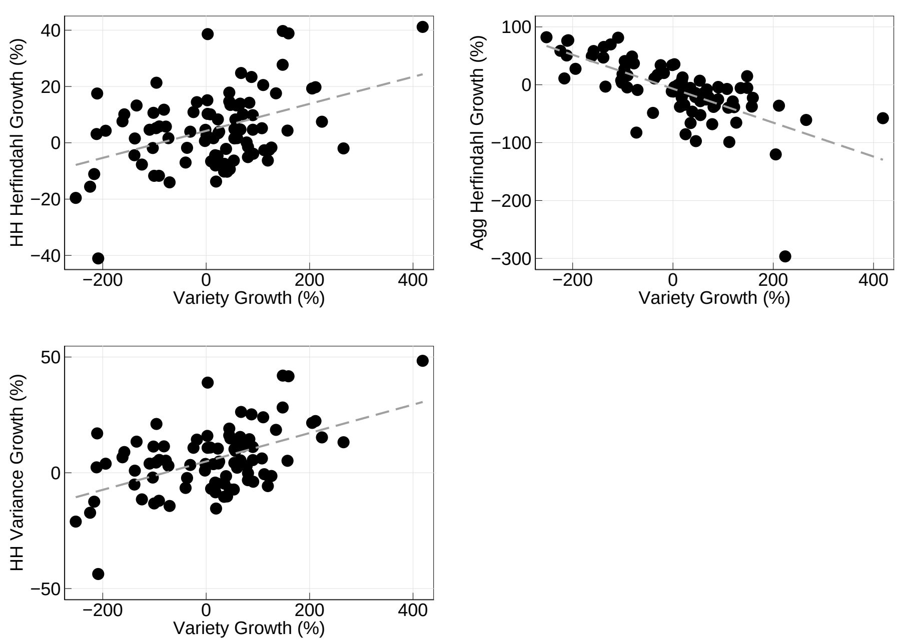
Figure 11: Within-Retailer Concentration and Variance



In particular, we recompute household concentration, variance and aggregate market shares, but computing market shares within a retailer × product group instead of just within a product group. Figure 11 shows that household concentration and cross-household product variance have both risen, even within retailers. Just like in the aggregate, variance has risen faster than household concentration so that aggregate concentration within the typical retailer has gone down. Thus, increasing segmentation and sorting across products has occurred even within retailers, suggesting that trends are not driven predominantly by the first explanation.

Finally, to more directly assess the extent to which expanded variety access may be leading to increased sorting, we explore the relationship between changes in concentration and household variance within a given retailer and the growth of variety availability at that same retailer. Consistent with an interpretation where increased availability of varieties is leading to increased household sorting, Table A.1 shows that this heterogeneity is strongly related to within retailer variety growth. The first three columns show results from the regression $\Delta_{2015-2004} \log outcome_i = \Delta_{2015-2004} \log varieties_i + \epsilon_i$ where $\Delta_{2015-2004} \log outcome_i$ is the growth of an outcome variable such as household Herfindahl for retailer i from 2004 to 2015 and $\Delta_{2015-2004} \log varieties_i$ is the growth of the total number of varieties consumed at that same retailer over the same period. We run this regression on a balanced panel of retailers in the sample from 2004-2015. Given the relatively small sample sizes, Figure 12 shows the scatter plot corresponding to these regression results. These results support the linear specification and show that statistical relationships are not driven by any particular outliers.

Figure 12: Retailer Variety Growth vs. Growth of Within-Retailer Concentration Statistics



In Columns 5-6 we instead run a panel regression with retailer fixed effects $\log outcome_i, t = \log varieties_i, t + \delta_i \epsilon_i, t$ using the unbalanced set of retailers. These results show that household concentration is significantly greater in retailers with greater variety growth. Higher variety growth also predicts an increase in cross-household variance and a decline in aggregate concentration.

8 Conclusions

Conventional measures of concentration and market shares study a producer's or product's share of aggregate spending. This paper instead poses that a different notion of "market" might result in more useful inference about changes in the same concepts of pricing power and consumer welfare that the conventional measures aimed to address. We unpack the consumer spending bundle and find that households – rich and poor, old and young – are increasingly concentrating their spending on particular products. While we cannot directly measure aggregate markups and market power, we do find a strong cross-household link between household concentration and effective market power: households with more concentrated purchases use coupons and exploit temporary sales less often and ultimately pay higher prices for the products they purchase. Further, households are concentrating their spending on different goods – superstar products aren't capturing more and more of our retail expenditures. Rather, as in many other walks of modern life, our grocery baskets look less and less like each otherâĂŹs. As a result, household spending bundles have become more concentrated at a time when aggregate spending on these same products has become less concentrated.

We associate these patterns with the introduction of new goods. Is product innovation the cause of increasing household segmentation or is it instead a response to more heterogeneous demand? [34] Is increasing polarization in other aspects of life driving such innovation or is it amplifying pre-existing differences across households? Does the answer to these question matter for inference about changes in aggregate market power and consumer welfare? We believe this paper opens the door to these

[34] We currently make no attempt to identify exogenous changes in product variety. See Jaravel (2017) for some interesting work using Bartik style variation to disentangle these effects in the context of household inflation differences.

intriguing questions, which we hope to address in our continuing work.

References

Aguado, J. M., C. Feijoo, and I. Martinez (2015): *Emerging Perspectives on the Mobile Content Evolution*. IGI Global.

Aguiar, M., and E. Hurst (2007): "Life-Cycle Prices and Production," *American Economic Review*, 97(5), 1533–1559.

Alwin, D., and P. Tufis (2015): "The Changing Dynamics of Class and Culture in American Politics: a Test of the Polarization Hypothesis," *The Annals of the American Academy of Political and Social Science*, 663.

Anderson, C. (2006): *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion.

Argente, D., and M. Lee (2017): "Cost of Living Inequality During the Great Recession," mimeo.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2017): "The Fall of the Labor Share and the rise of Superstar Firms," Working Paper 23396, National Bureau of Economic Research.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006): "The Polarization of the U.S. Labor Market," *The American Economic Review*, 96(2).

Barkai, S. (2017): "Declining Labor and Capital Shares," mimeo.

Berger, D., and J. Vavra (2018): "Shocks vs. Responsiveness: What Drives Time-Varying Dispersion?," *Journal of Political Economy*.

Bertrand, M., and E. Kamenica (2018): "Coming apart? Lives of the Rich and the Poor Over Time," Discussion paper.

De Loecker, J., and J. Eeckhout (2017): "The Rise of Market Power and the Macroeconomic Implications," UCL mimeo.

Dhyne, E., A. Kikkawa, and G. Magerman (2017): "Imperfect Competition and the Transmission of Shocks: The Network Matters," .

Dube, J.-P., G. Hitsch, and P. Rossi (2018): "Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession," *Marketing Science*.

GENTZKOW, M., J. M. SHAPIRO, AND M. TADDY (2017): "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech," Working Paper 22423, National Bureau of Economic Research.

HANSEN, K., AND V. SINGH (2015): "Choice Concentration," Discussion paper.

JARAVEL, X. (2017): "The Unequal Gains from Product Innovations: Evidence from the US Retail Sector," mimeo.

KARABARBOUNIS, L., AND B. NEIMAN (2018): "Accounting for Factorless Income," Working Paper.

OLIVIER COIBION, YURIY GORODNICHENKO, D. K. (2017): "Consumption Inequality and the Frequency of Purchases," Discussion paper.

PEW RESEARCH CENTER (2014): "Political Polarization in the American Public," Discussion paper.

PIKETTY, T., E. SAEZ, AND G. ZUCMAN (2016): "Distributional National Accounts: Methods and Estimates for the United States," Working Paper 22945, National Bureau of Economic Research.

RADAELLI, P., AND M. ZENGA (2002): "A New Two Term Subtractive Decomposition Of Herfindahl Concentration Measure," in *XLI Riunione Scientifica della Società Italiana di Statistica*, pp. 525–528. Società Italiana di Statistica.

STROEBEL, J., AND J. VAVRA (2018): "House Prices, Local Demand and Retail Prices," *Journal of Political Economy*.

TRAINA, J. (2018): "Markup Estimation Using Financial Statements: Cost of Goods Sold vs. Operating Expense," University of Chicago Working Paper.

Table 1: Demographics and Household Concentration Levels

	(1) $H_{i,c}^{t}$	(2) $H_{i,c}^{t}$	(3) $H_{i,c}^{t}$	(4) $H_{i,c}^{t}$	(5) $H_{i,c}^{t}$
household size 2	-0.045***	-0.044***	-0.044***	-0.033***	-0.032***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
household size 3	-0.067***	-0.062***	-0.062***	-0.049***	-0.048***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
household size 4	-0.077***	-0.070***	-0.070***	-0.052***	-0.051***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
household size 5+	-0.084***	-0.077***	-0.077***	-0.054***	-0.054***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
age 30-34		-0.002	-0.002	0.000	-0.000
		(0.002)	(0.002)	(0.002)	(0.002)
age 35-39		-0.003	-0.003	0.002	0.002
		(0.002)	(0.002)	(0.002)	(0.002)
age 40-44		-0.002	-0.002	0.008***	0.008***
		(0.003)	(0.003)	(0.003)	(0.003)
age 45-49		-0.000	-0.000	0.013***	0.014***
		(0.004)	(0.004)	(0.004)	(0.003)
age 50-54		0.005	0.005	0.019***	0.020***
		(0.005)	(0.005)	(0.004)	(0.004)
age 55-59		0.009	0.009	0.025***	0.026***
		(0.006)	(0.006)	(0.005)	(0.004)
age 60-64		0.011*	0.011*	0.029***	0.030***
		(0.006)	(0.006)	(0.005)	(0.005)
age 65-69		0.010	0.010	0.027***	0.028***
		(0.006)	(0.006)	(0.005)	(0.005)
age 70+		0.018**	0.018**	0.031***	0.030***
		(0.007)	(0.007)	(0.006)	(0.005)
Income $20000-34999			-0.002	-0.000	-0.001
			(0.003)	(0.002)	(0.002)
Income $35000-59999			-0.003	0.001	-0.000
			(0.004)	(0.004)	(0.003)
Income >$60000			-0.002	0.008	0.004
			(0.005)	(0.005)	(0.004)
log spend				-0.064***	-0.064***
				(0.006)	(0.006)
Product Group Fixed Effect	YES	YES	YES	YES	YES
Additional Controls	NO	NO	NO	NO	YES
R-squared	0.320	0.321	0.321	0.379	0.380
N	45,476,529	45,476,529	45,476,529	45,476,529	45,476,162

Note: Table shows results from a regression of household-product group-year herfindahls on various demographic variables plus a product group fixed effect. Omitted categorical variables are household size 1, age<=29, and income<$20000. In column (5) additional controls are: dummy variables for education, scantrack markets, type of residence, and race. The unit of observation is a household-product group-year, and observations are weighted using Nielsen sampling weights × household-year-product group spending. Standard errors shown in parantheses are two-way clustered by household and product group. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 2: Effect of Demographics on Concentration Trends

	(1)	(2)	(3)	(4)	(5)	(6)
	$H^t_{i,c}$	$H^t_{i,c}$	$H^t_{i,c}$	$H^t_{i,c}$	$H^t_{i,c}$	$H^t_{i,c}$
t/100	0.272***	0.258***	0.250***	0.250***	0.349***	0.350***
	(0.042)	(0.041)	(0.041)	(0.041)	(0.038)	(0.038)
household size 2		-0.045***	-0.044***	-0.044***	-0.032***	-0.031***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
household size 3		-0.066***	-0.062***	-0.062***	-0.048***	-0.047***
		(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
household size 4		-0.076***	-0.070***	-0.070***	-0.052***	-0.051***
		(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
household size 5		-0.084***	-0.077***	-0.077***	-0.054***	-0.054***
		(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
age 30-34			-0.002	-0.002	0.001	0.000
			(0.002)	(0.002)	(0.002)	(0.002)
age 35-39			-0.003	-0.003	0.003	0.003
			(0.002)	(0.002)	(0.002)	(0.002)
age 40-44			-0.001	-0.001	0.009***	0.010***
			(0.003)	(0.003)	(0.003)	(0.003)
age 45-49			0.000	0.000	0.014***	0.015***
			(0.004)	(0.004)	(0.003)	(0.003)
age 50-54			0.004	0.004	0.018***	0.019***
			(0.005)	(0.005)	(0.004)	(0.004)
age 55-59			0.008	0.008	0.023***	0.024***
			(0.005)	(0.006)	(0.005)	(0.004)
age 60-64			0.009	0.009	0.027***	0.028***
			(0.006)	(0.006)	(0.005)	(0.005)
age 65-69			0.009	0.009	0.025***	0.026***
			(0.006)	(0.006)	(0.005)	(0.005)
age 70+			0.018**	0.018**	0.030***	0.030***
			(0.007)	(0.007)	(0.006)	(0.005)
Income $20000-34999				-0.002	0.000	-0.000
				(0.003)	(0.002)	(0.002)
Income $35000-59999				-0.003	0.002	0.000
				(0.004)	(0.004)	(0.003)
Income >$60000				-0.002	0.008	0.003
				(0.005)	(0.005)	(0.004)
log spend					-0.064***	-0.065***
					(0.006)	(0.006)
Product Group Fixed Effect	YES	YES	YES	YES	YES	YES
Additional Controls	NO	NO	NO	NO	NO	YES
R-squared	0.310	0.321	0.322	0.322	0.381	0.382
N	45,476,529	45,476,529	45,476,529	45,476,529	45,476,529	45,476,162

Note: Table shows results from a regression of household-product group-year herfindahls on a time trend, various demographic variables plus and product group fixed effect. Omitted categorical variables are household size 1, age<=29, and income<$20000. In column (5) additional controls are: dummy variables for education, scantrack markets, type of residence, and race. The unit of observation is a household-product group-year, and observations are weighted using Nielsen sampling weights × household-year-product group spending. Standard errors shown in parantheses are two-way clustered by household and product group. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 3: Concentration trends by demographic groups

	(1) $H^t_{i,c}$	(2) $H^t_{i,c}$	(3) $H^t_{i,c}$	(4) $H^t_{i,c}$
household size 1 × t/100	0.346*** (0.049)			
household size 2 × t/100	0.339*** (0.040)			
household size 3 × t/100	0.365*** (0.041)			
household size 4 × t/100	0.371*** (0.038)			
household size 5 × t/100	0.339*** (0.044)			
age <30 × t/100		0.393*** (0.053)		
age 30-34 × t/100		0.370*** (0.044)		
age 35-39 × t/100		0.347*** (0.038)		
age 40-44 × t/100		0.348*** (0.042)		
age 45-49 × t/100		0.322*** (0.043)		
age 50-54 × t/100		0.312*** (0.045)		
age 55-59 × t/100		0.371*** (0.045)		
age 60-64 × t/100		0.334*** (0.057)		
age 65-69 × t/100		0.419*** (0.047)		
age 70+ × t/100		0.348*** (0.048)		
Income < $20000 × t/100			0.398*** (0.058)	
Income $20000-34999 × t/100			0.354*** (0.044)	
Income $35000-59999 × t/100			0.351*** (0.037)	
Income >$60000 × t/100			0.334*** (0.037)	
< HS × t/100				0.474*** (0.162)
Some HS × t/100				0.231*** (0.085)
HS × t/100				0.359*** (0.045)
Some College × t/100				0.342*** (0.039)
College × t/100				0.354*** (0.036)
Post College × t/100				0.360*** (0.041)
Product Group Fixed Effect	YES	YES	YES	YES
Additional Controls	YES	YES	YES	YES
R-squared	0.382	0.382	0.382	0.382
N	45,476,162	45,476,162	45,476,162	45,476,162

Note: Table shows results from a regression of household-product group-year herfindahls on a time trend interacted with different demographic groups, various demographic variables controls in levels and product group fixed effect. All regressions include additional controls for log spending and dummy variables for education, scantrack markets, type of residence, and race. The unit of observation is a household-product group-year, and observations are weighted using Nielsen sampling weights × household-year-product group spending. Standard errors shown in parantheses are two-way clustered by household and product group. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

Table 4: Household Concentration and Relative Prices

	(1)	(2)	(3)	(4)
	$log\hat{P}_i^t$	$log\hat{P}_i^t$	$log\hat{P}_i^t$	$log\hat{P}_i^t$
$\overline{H}_i^t$	0.116***	0.162***	0.055***	0.103***
	(0.004)	(0.004)	(0.003)	(0.004)
Household FE	NO	NO	YES	YES
Additional Controls	NO	YES	NO	YES
R-squared	0.019	0.112	0.842	0.846
N	668779	664515	668779	664515

Note: Table shows results from a regression of household log relative price indices on household herfindahls. Columns (3) and (4) include household FE and so are identified only off within-household changes in concentration and relative prices and not off of persistent level differences. In columns (2) and (4) additional controls are: a time trend, log household spending, plus dummy variables for household size, age groups, income groups, for education, employment status, occupation, scantrack markets, marital status, type of residence, race, presence of children, presence of household internet, cable/non-cable tv, and various indicators for the presence of major kitchen appliances. The unit of observation is a household-year, and observations are weighted using Nielsen sampling weights. Standard errors shown in parantheses are clustered by household. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 5: Household Concentration and Coupon and Sales Intensity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Frac Coupon	Frac Sale	Frac Coupon	Frac Sale	Diff Coupon	Diff Sale	Diff Coupon	Diff Sale
$\overline{H}_i^t$	-0.124***	-0.275***	-0.132***	-0.398***	-0.049***	-0.084***	-0.052***	-0.170***
	(0.006)	(0.010)	(0.006)	(0.012)	(0.005)	(0.009)	(0.005)	(0.010)
Additional Controls	NO	NO	YES	YES	NO	NO	YES	YES
R-squared	0.009	0.014	0.068	0.103	0.002	0.002	0.025	0.059
N	668779	668779	664515	664515	668779	668779	664515	664515

Note: First four columns show a regression of the fraction of household purchases using coupons and on sale on household concentration. The outcome variable in columns (5)-(8) is the difference between the fraction of purchases for a household's top product which are made using coupons or on sale and the average fraction for that same product across all households. Columns (3) and (4) and (7) and (8) include additional controls for a time trend, log household spending, plus dummy variables for household size, age groups, income groups, for education, employment status, occupation, scantrack markets, marital status, type of residence, race, presence of children, presence of household internet, cable/non-cable tv, and various indicators for the presence of major kitchen appliances. The unit of observation is a household-year, and observations are weighted using Nielsen sampling weights. Standard errors shown in parantheses are clustered by household. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 6: Effect of Coupon and Sales Intensity on Concentration-Price Relationship

	(1) $log\hat{P}_i^t$	(2) $log\hat{P}_i^t$	(3) $log\hat{P}_i^t$	(4) $log\hat{P}_i^t$
$\overline{H}_i^t$	0.116***	0.062***	0.162***	0.098***
	(0.004)	(0.003)	(0.004)	(0.003)
Frac Coupon		-0.410***		-0.408***
		(0.004)		(0.004)
Frac Sale		-0.013***		-0.023***
		(0.002)		(0.002)
Additional Controls	NO	NO	YES	YES
R-squared	0.019	0.427	0.112	0.513
N	668779	668779	664515	664515

Note: Table shows results from a regression of household log relative price indices on household herfindahls with and without controls for household coupon and sales intensity. Additional controls are: a time trend, log household spending, plus dummy variables for household size, age groups, income groups, for education, employment status, occupation, scantrack markets, marital status, type of residence, race, presence of children, presence of household internet, cable/non-cable tv, and various indicators for the presence of major kitchen appliances. The unit of observation is a household-year, and observations are weighted using Nielsen sampling weights. Standard errors shown in parantheses are clustered by household. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 7: Within-Retailer Variety Growth vs Concentration Trends

	(1) $\Delta \log \overline{H}^t$	(2) $\Delta \log H_{agg}^t$	(3) $\Delta \log \overline{V}^t$	(4) $log\overline{H}^t$	(5) $\log H_{agg}^t$	(6) $\log \overline{V}^t$
$\Delta \log varieties$	0.048***	-0.293***	0.061***			
	(0.013)	(0.052)	(0.013)			
$\log varieties$				0.036***	-0.238***	0.049***
				(0.013)	(0.027)	(0.014)
Retailer FE	NO	NO	NO	YES	YES	YES
R-squared	0.172	0.444	0.252	0.960	0.909	0.959
N	88	88	88	1619	1608	1619

Note: Table shows results from a regression of within retailer household concentration, “aggregate" concentration, and household variance on retailer variety counts. Columns 1-3 are long difference specifications from 2004-2015 using a balanced panel of retailers. Columns 4-6 are a panel regression with retailer fixed effects. Standard errors shown in parantheses are clustered by retailer. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

A Appendix

Figure A.1: Average number of UPCs per retailer-product group in Nielsen Retailer RMS data, weighted by retailer-product group spending

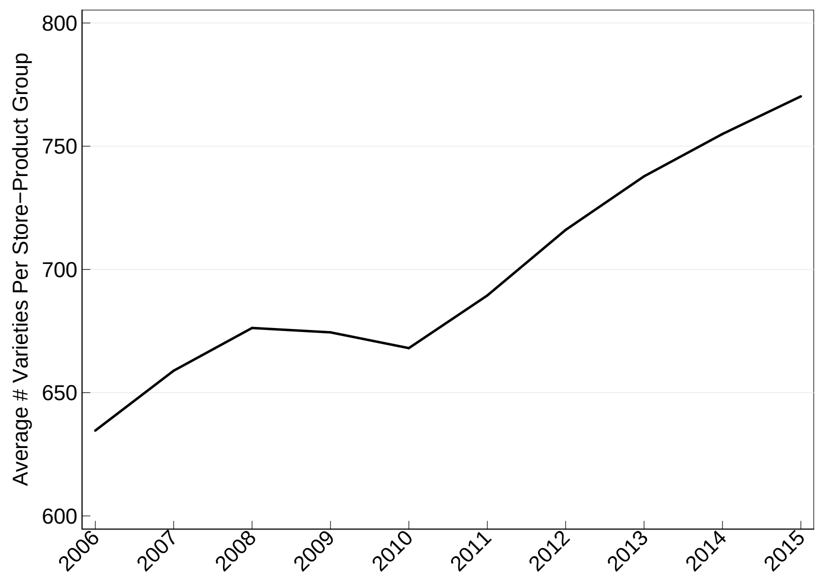


Figure A.2: Household Product (UPC) Concentration (Including Generics)

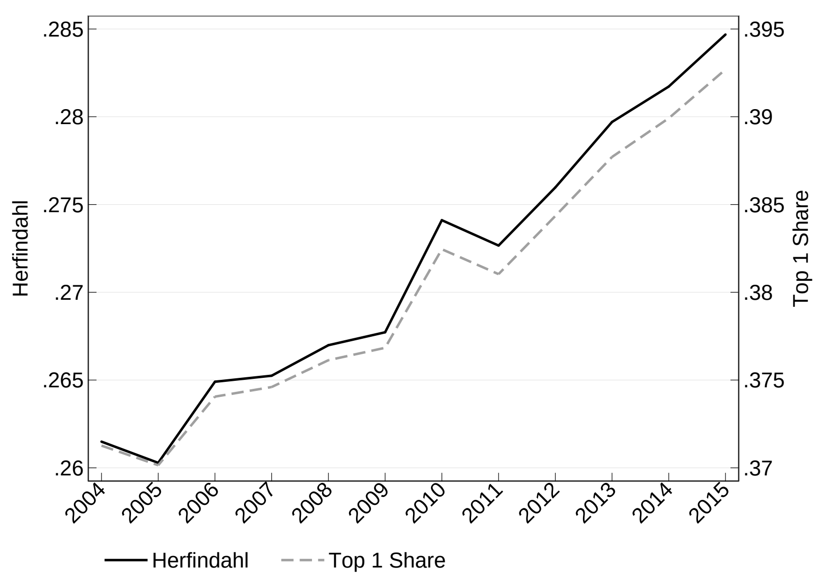


Figure A.3: Household Product (Brand) Concentration

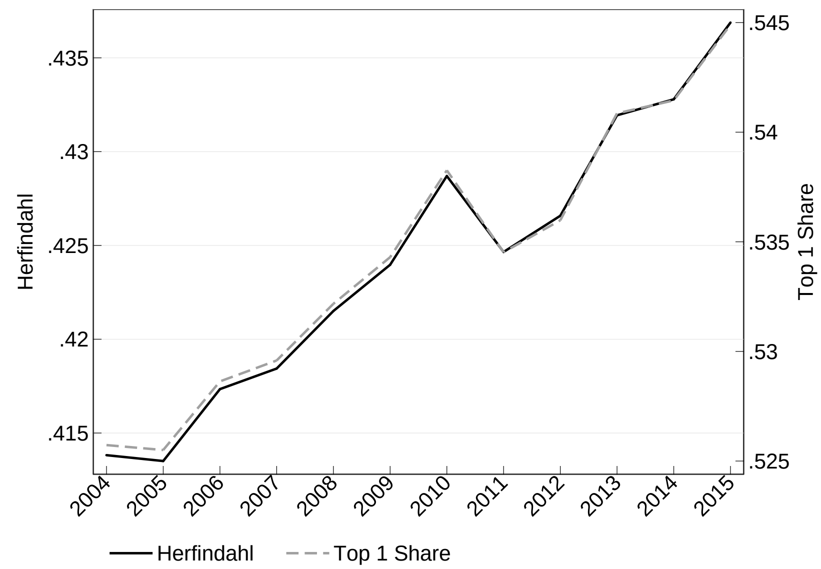


Figure A.4: Household Product (UPC) Concentration: Including Category Composition Changes

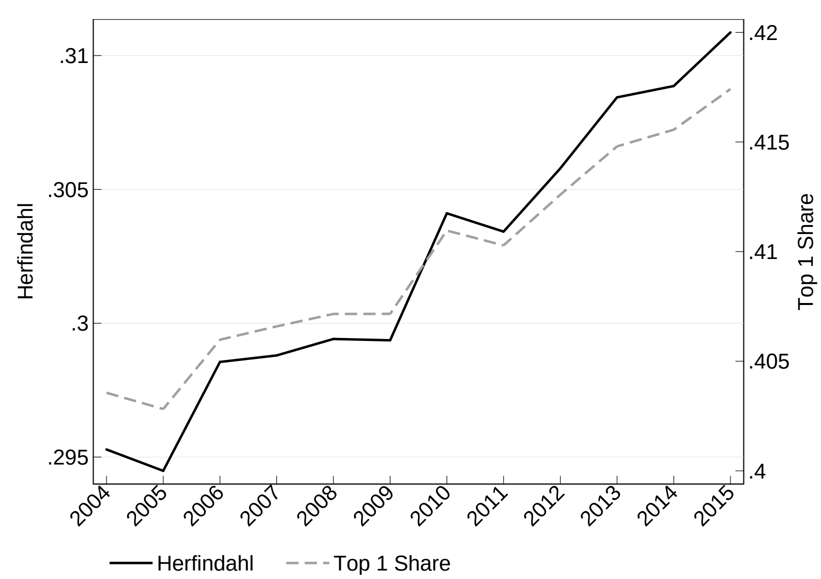
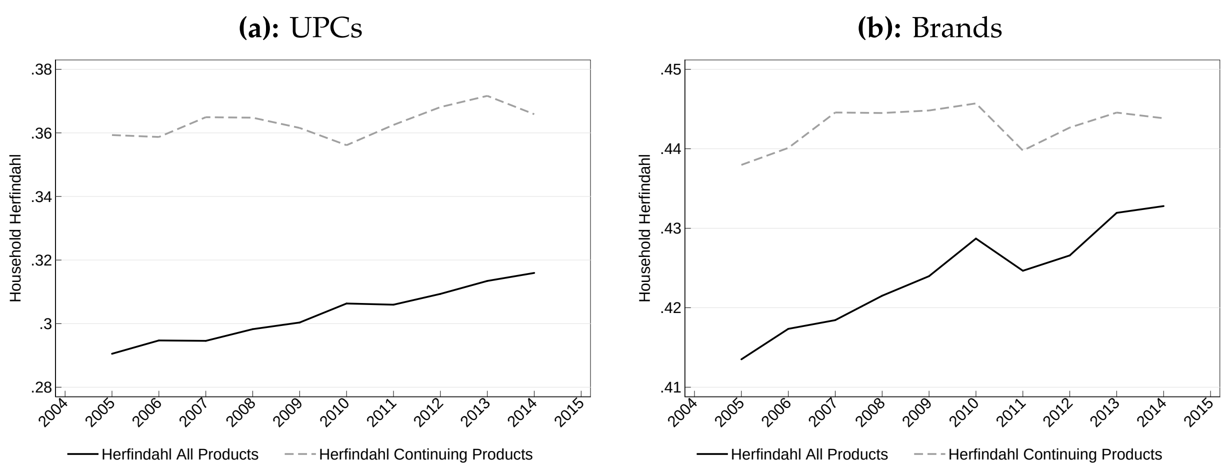


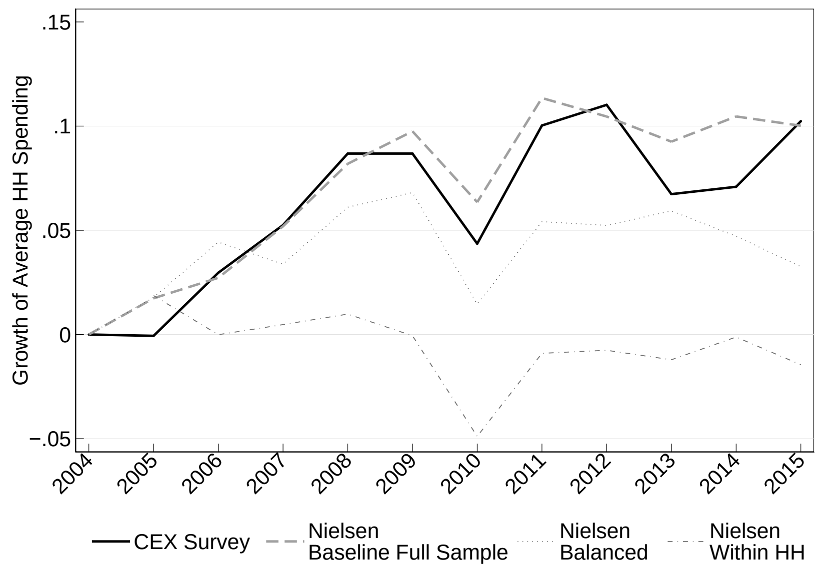
Figure A.5: 2004-2015 Household Herfindahl growth for continuing products (defining “new" at aggregate rather than household level)



B Appendix on measurement error/attrition bias

How well does Nielsen spending line up with external measures from the Consumer Expenditure Survey on similar categories?

Figure A.6: Household Spending in Nielsen vs. Consumer Expenditure Survey

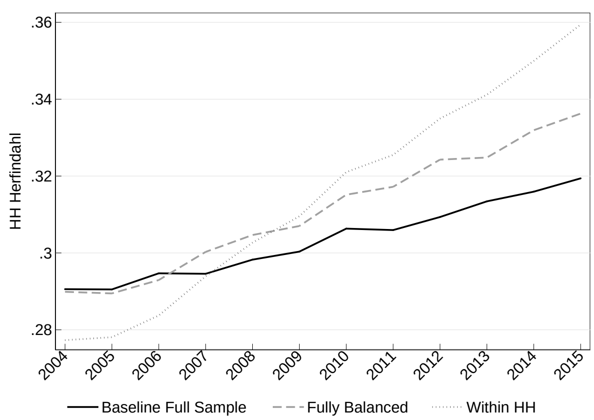


Why is within-household spending growth substantially less strong than overall household spending? Likely two reasons: 1) The panel dimension of Nielsen is not representative of all households. The continuing households in the sample are substantially older than the overall Nielsen sample and the overall population, and we know from other research that households around retirement have declining food spending. While Nielsen provides sampling weights to make the overall sample representative of the U.S., there are no provided weights to reweight the panel dimension to be representative of the overall US, and the requisite demographic variables in the data to construct them ourselves do not exist. 2) There is likely attrition bias and households probably report a declining share of spending across time. This attrition bias may be particularly strong in the final year in which a household is in the sample, which could potentially explain the difference between the balanced and within household lines. If reduced reporting tends to proceed exit, then one would expect attrition bias to be less severe for households who remain in the sample for the full 12 years. Consistent with this, the balanced sample exhibits stronger spending growth than the within household sample.

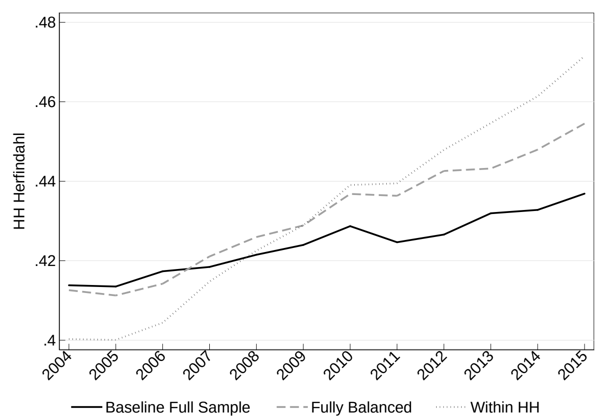
For these reasons, our baseline results use the entire Nielsen homescan panel rather than focusing on a balanced panel of households. However, it is useful to compare our basic trends in the full sample to those computed using within-household variation. Figure A.7 thus redoes Figure 1 using a fully balanced panel and with a specification using only the within-household changes specification.

Figure A.7: Household Herfindahl Trends for Different Samples

(a): UPC



(b): Brand



Clearly trends are even stronger than our baseline results when using the fully balanced panel or when identifying off of within-household variation, so in this sense our baseline is conservative. We now describe several forces that might spuriously increase the within-household trend as well as some alternative forces which might spuriously flatten the full sample trend. This makes it difficult to know whether our baseline sample is likely to be understated or whether it is instead the balanced panel specification that is overstated. However, in either case, the trend is robustly positive, and our baseline sample is the one which generates more conservative results.

More specifically, the full sample trend could potentially be biased downwards because the Nielsen sampling technology changes across time, and these changes are implemented when households enter the sample. These changes in technology could obscure underlying trends in the data, but would be stripped out when using within-household variation. More generally, households have very different concentration levels, as shown above, so that random household entry and exit in the sample could make it more difficult to pick up underlying trends. These are both forces that might lead our baseline full sample to understate the true increase in concentration across time.

Conversely, we have shown above both that increases in spending are strongly negatively correlated with increases in concentration and that the within-household sample has spending growth much lower than in the consume expenditure survey. To the extent that the within-household sample has spurious declining spending due to sample attrition, there is then a concern that using within household variation might lead to an upward biased trend. However, if we redo all our regression results using within household variation *controlling* for within household changes in spending, we continue to find upward trends which are stronger than in the full sample. This suggests that the stronger upward trend in the within-household results is not driven solely by the lower reported spending growth in this sample. In addition, we can also recompute results using only households in the first year in the sample. By construction, attrition bias in spending across time cannot drive any trend, since this sample has no within-household time-series variation but it still delivers an upward trend. Finally, attrition bias is less likely to be a concern for the fully balanced sample: The upward trend in the fully balanced panel is roughly linear across time, so if this upward trend was explained by attrition bias and progressive under reporting, this under reporting would need to grow at a constant rate, which seems unlikely, especially because Nielsen tries to drop households from the sample who are not reporting accurately. It seems much more likely that the biggest under reporting would occur in the first year or two in the panel as households are likely to be most enthusiastic about scanning purchases initially and then reduce scanning as it becomes more tedious across time. It would be quite surprising if enthusiasm waned at a constant linear rate across time but that households continued to participate in the homescan panel.

Together, we think that these results suggest the stronger upward trends using the balanced samples and the within-sample variation are not driven by spurious attrition bias. Nevertheless, we cannot fully rule out this concern. Furthermore, as discussed above the panel element of the sample is not representative since households who remain in the sample for progressive years are demographically different and not representative of the population leading total spending for this population to line up less well with aggregate spending inferred from the consumer expenditure survey. For these reasons and to be conservative, we focus on the full sample in all our baseline results but only note here that using other samples only strengthens our conclusions.

C Appendix Tables

Table A.1: Demographics and Household Concentration Levels: Household Level Regression, no Product Group Composition Controls

	(1) $\overline{H}_i^t$	(2) $\overline{H}_i^t$	(3) $\overline{H}_i^t$	(4) $\overline{H}_i^t$	(5) $\overline{H}_i^t$
household size 2	-0.063***	-0.062***	-0.062***	-0.020***	-0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
household size 3	-0.091***	-0.084***	-0.084***	-0.037***	-0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
household size 4	-0.108***	-0.098***	-0.098***	-0.044***	-0.028***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
household size 5+	-0.116***	-0.106***	-0.106***	-0.046***	-0.031***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
age 30-34		-0.002	-0.003	0.002	0.002
		(0.002)	(0.002)	(0.001)	(0.001)
age 35-39		-0.002	-0.002	0.007***	0.005***
		(0.002)	(0.002)	(0.001)	(0.001)
age 40-44		0.002	0.001	0.015***	0.012***
		(0.002)	(0.002)	(0.001)	(0.001)
age 45-49		0.006***	0.005***	0.024***	0.020***
		(0.002)	(0.002)	(0.001)	(0.001)
age 50-54		0.014***	0.013***	0.033***	0.027***
		(0.002)	(0.002)	(0.001)	(0.001)
age 55-59		0.020***	0.019***	0.039***	0.032***
		(0.002)	(0.002)	(0.001)	(0.002)
age 60-64		0.023***	0.022***	0.045***	0.037***
		(0.002)	(0.002)	(0.002)	(0.002)
age 65-69		0.021***	0.021***	0.044***	0.036***
		(0.002)	(0.002)	(0.002)	(0.002)
age 70+		0.024***	0.024***	0.037***	0.031***
		(0.002)	(0.002)	(0.002)	(0.002)
Income \$20000-34999			-0.005***	0.003***	0.001
			(0.001)	(0.001)	(0.001)
Income \$35000-59999			-0.005***	0.009***	0.004***
			(0.001)	(0.001)	(0.001)
Income >\$60000			0.001	0.023***	0.013***
			(0.001)	(0.001)	(0.001)
log *spend*				-0.094***	-0.093***
				(0.001)	(0.001)
Additional Controls	NO	NO	NO	NO	YES
R-squared	0.130	0.136	0.137	0.361	0.381
N	668779	668779	668779	668779	664515

Note: Table shows results from a regression of household herfindahls on various demographic variables. Omitted categorical variables are household size 1, age<=29, and income<\$20000. In column (5) additional controls are: dummy variables for education, employment status, occupation, scantrack markets, marital status, type of residence, race, presence of children, presence of household internet, cable/non-cable tv, and various indicators for the presence of major kitchen appliances. The unit of observation is a household-year, and observations are weighted using Nielsen sampling weights. Standard errors shown in parantheses are clustered by household. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table A.2: Effect of Demographics on Concentration Trends: Household Level Regression, no Product Group Composition Controls

	(1) $\overline{H}_i^t$	(2) $\overline{H}_i^t$	(3) $\overline{H}_i^t$	(4) $\overline{H}_i^t$	(5) $\overline{H}_i^t$	(6) $\overline{H}_i^t$
t	0.233***	0.226***	0.217***	0.215***	0.223***	0.194***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.009)
household size 2		-0.063***	-0.062***	-0.062***	-0.020***	-0.015***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
household size 3		-0.091***	-0.084***	-0.084***	-0.036***	-0.023***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
household size 4		-0.108***	-0.098***	-0.098***	-0.044***	-0.028***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
household size 5+		-0.116***	-0.106***	-0.106***	-0.046***	-0.031***
		(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
age 30-34			-0.002	-0.002	0.003*	0.002
			(0.002)	(0.002)	(0.001)	(0.001)
age 35-39			-0.001	-0.002	0.008***	0.006***
			(0.002)	(0.002)	(0.001)	(0.001)
age 40-44			0.003*	0.002	0.017***	0.013***
			(0.002)	(0.002)	(0.001)	(0.001)
age 45-49			0.007***	0.006***	0.025***	0.020***
			(0.002)	(0.002)	(0.001)	(0.001)
age 50-54			0.014***	0.013***	0.033***	0.027***
			(0.002)	(0.002)	(0.001)	(0.001)
age 55-59			0.019***	0.019***	0.039***	0.032***
			(0.002)	(0.002)	(0.001)	(0.002)
age 60-64			0.022***	0.022***	0.044***	0.036***
			(0.002)	(0.002)	(0.002)	(0.002)
age 65-69			0.021***	0.021***	0.043***	0.035***
			(0.002)	(0.002)	(0.002)	(0.002)
age 70+			0.025***	0.025***	0.038***	0.030***
			(0.002)	(0.002)	(0.002)	(0.002)
Income $20000-34999				-0.004***	0.003***	0.001
				(0.001)	(0.001)	(0.001)
Income $35000-59999				-0.005***	0.009***	0.004***
				(0.001)	(0.001)	(0.001)
Income >$60000				-0.000	0.022***	0.013***
				(0.001)	(0.001)	(0.001)
ls					-0.094***	-0.093***
					(0.001)	(0.001)
Additional Controls	NO	NO	NO	NO	NO	YES
R-squared	0.005	0.135	0.140	0.141	0.365	0.383
N	668779	668779	668779	668779	668779	664515

Note: Table shows results from a regression of household herfindahls on various demographic variables and a time trend. Omitted categorical variables are household size 1, age<=29, and income<$20000. In column (5) additional controls are: dummy variables for education, employment status, occupation, scantrack markets, marital status, type of residence, race, presence of children, presence of household internet, cable/non-cable tv, and various indicators for the presence of major kitchen appliances. The unit of observation is a household-year, and observations are weighted using Nielsen sampling weights. Standard errors shown in parantheses are clustered by household. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).