

The Rise of Niche Consumption*

Brent Neiman
University of Chicago

Joseph Vavra
University of Chicago

December 2019

Abstract

We show that over the last 15 years, the typical household has increasingly concentrated its spending on a few preferred products. However, this is not driven by “superstar” products capturing larger market shares. Instead, households increasingly focus spending on different products from each other. As a result, aggregate spending concentration has in fact decreased over this same period. We use a novel heterogeneous agent model to conclude that increasing product variety is a key driver of these divergent trends. When more products are available, households can select a subset better matched to their particular tastes, and this generates welfare gains not reflected in government statistics. Our model features heterogeneous markups because producers of popular products care more about maximizing profits from existing customers, while producers of less popular niche products care more about expanding their customer base. Surprisingly, however, our model can match the observed trends in household and aggregate concentration without any resulting change in aggregate market power.

JEL-Codes: E21, E31, D12, D4

Keywords: Product Concentration, Niche Products, Market Power, Markups, Long-tail

*We thank David Argente, Anhua Chen, Levi Crews, and Agustin Gutierrez for providing exceptional research assistance and Rodrigo Adao, Jonathan Dingel, J.P. Dubé, Austan Goolsbee, Pete Klenow, Thomas Mertens, Esteban Rossi-Hansberg, and Tom Wollmann for helpful comments and suggestions. Our analyses are calculated or derived based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at 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. This paper supplants a previous draft circulated as “The Rise of Household Spending Concentration.”

1 Introduction

We show that households are increasingly buying different products from each other. Using a novel model of consumers with heterogeneous tastes, we argue that expanding product availability is crucial for understanding this phenomenon. As new products are introduced, households can choose consumption bundles better suited to their particular tastes, resulting in welfare gains from better selection. Fitting our model to detailed micro data on heterogeneous household spending, we identify and quantify the growth of product availability and associated welfare gains in more than one hundred separate product categories, covering a large fraction of consumer spending on groceries and other household nondurables. We find large increases in product availability and resulting welfare gains from selection in most categories that cannot be identified with standard representative agent macro models.

We start by documenting that since 2004, the typical household has increased the concentration of its spending on a few preferred products. This fact on its own might point toward an increasing importance of “superstar” products, but a similar analysis of aggregate spending paints a different picture. Pooling households together, we see that total spending on these same products over the same period has in fact become more evenly distributed. That is, aggregate concentration has declined despite the increase in household concentration. This divergence implies that while each household increasingly concentrates spending on its own preferred products, households also increasingly differ on which products they consume. We refer to this greater fragmentation of the product space as a rise in “niche” consumption.

The rise in niche consumption occurs fairly steadily over time and is robust to a variety of specification and measurement choices as well as to the inclusion of a variety of controls for observables. Interestingly, the divergence between household and aggregate concentration is not driven by a widening gap between the rich and poor households, between consumers in one region and another, or by differences between households grouped according to other demographic characteristics. Rather, we find that household consumption bundles are becoming more differentiated even when measured within cities, within store chains, and within demographic groups defined by income, race, education, age, and household size. Niche consumption also grows in almost all product categories.

What then drives these divergent trends, and should we care about them? Many standard models cannot be used to answer these questions since they rule out the differential trends in household and aggregate concentration by assumption. For instance, any representative household model will exhibit identical household and aggregate concentration. Standard discrete choice models imply that household spending within categories is completely concentrated on a single product.¹ Instead, we

¹Dynamic discrete choice models with temporal aggregation could likely also speak to our primary empirical facts, which focus on annual household spending. However, even when looking at individual shopping trips made by one-person households, it is common for multiple products to be simultaneously purchased in a single category. See Appendix C.

build a model of consumers that exhibit a love-of-variety and where their preference ordering across products differs from each other.

Our model of an individual household follows Li (2018) and features constant elasticity of substitution (CES) preferences, product-specific taste shocks, and a utility cost borne per variety consumed. Under these assumptions, the household chooses to consume only a subset of the total available products. When tastes for products, adjusted for prices, are distributed Pareto, we obtain a closed-form expression relating the household's Herfindahl index – the concentration measure we use in our empirical analyses – to structural parameters of the model.

To move from an analysis of an individual household to an analysis of aggregate spending, we introduce a continuum of households that are identical except for the ordering of their preferences across products. We assume that all households have tastes that decline identically from their favorite product to their second favorite, and so on, so all households have identical taste *distributions*. However, the actual identity of these first- and second-favorite products differ from one household to the next.

We introduce a “rank” function, which maps each product to a relative position in each household's tastes. A household's rank for a given product is a weighted average of that product's aggregate component, which is common across all households, and a random household-specific component. If the aggregate component receives all the weight, the environment collapses to a representative household economy with all households consuming the same products and with equal household and aggregate spending concentration. Conversely, if the household-specific component receives all the weight, there will be uniform aggregate spending across products and low aggregate concentration, even if individual household spending is highly concentrated. We analyze an empirically-disciplined intermediate case and obtain another closed-form expression relating the aggregate Herfindahl to structural parameters in the model.

Interestingly, we next show that in this intermediate case, different products in the economy face different elasticities of demand even though household preferences are CES. This is because when a product reduces its price, it expands sales to existing customers with a constant elasticity, but it also attracts new customers who previously did not purchase the product at all. The relative strength of these forces varies with products' market shares, which generates variable elasticities of demand and implied markups. Our model assumptions allow us to characterize the distribution of elasticities as well as the implied aggregate markup in the economy analytically.

The closed-form analytical solutions for various aggregates that arise from our particular parametric assumptions in turn deliver precise conclusions about identification. This is important because key structural parameters including the elasticity of substitution, the utility costs of variety, and the shape of the taste distribution are all either unobserved or are challenging to estimate directly in the data. Even the total number of available products, a key input into our model, cannot be confidently counted in the Nielsen data as its level and trend are highly sensitive to the choice to include or exclude prod-

ucts accounting for tiny amounts of aggregate spending.² Therefore, we use the model’s analytical expressions, together with key moments from the data, to compute changes in the value of structural parameters and assess the implications of these changes for welfare.

We find that even when we allow for simultaneous movements in all structural parameters, matching the observed divergence between household and aggregate concentration can *only* be achieved with sizable growth in the number of varieties, of roughly 5 percent per year. This increase in available varieties leads households to endogenously consume more products and enjoy “love-of-variety” welfare gains, as is standard in CES environments. In our model, an increase in the total number of varieties available also allows households to select a subset of products better matched to their particular tastes and thus enjoy what we call “gains from selection”.

The strength of these two effects depends on the elasticity of substitution and the shape parameter governing the asymmetry of tastes, which can be identified given data on aggregate markups. For plausible targets for the aggregate markup, we find that the increase in product availability necessary to hit the divergence between household and aggregate concentration leads to large welfare gains, which are mostly driven by the gains from selection. For example, if we hold all other parameters fixed in our preferred calibration and introduce 5 percent annual growth in the number of available products into our model, we calculate that consumption-equivalent welfare grows by 0.76 percent per year, and that 89 percent of this growth comes from gains to selection.³ Since these gains from selection need not be captured by typical matched-model price indices used by national statistical agencies, our model reveals how expanding product availability in an environment with heterogeneous tastes may generate significant unmeasured gains in standards of living.

While an increase in variety availability is necessary to generate the divergence between household and aggregate concentration, the model implies that this increase is not sufficient to match all the empirical trends we document. In particular, an increase in variety availability, on its own, will actually cause household concentration to fall mildly, even though aggregate concentration will fall by much more. We therefore use the model to infer additional changes in the fixed cost per variety, the elasticity of substitution and the shape of the taste distribution to exactly both the observed trends in household and aggregate Herfindahls, as well as the relationship between the household Herfindahl and the number of varieties consumed by individual households.

We emphasize that, unlike the increase in the number of available varieties, changes in these other structural parameters are not fully identified and various configurations are consistent with the data.⁴ However, we consider various combinations of parameter changes in addition to the increase in avail-

²See Appendix C for extensive discussion.

³Our preferred calibration targets an aggregate markup (ratio of aggregate sales to aggregate costs) of 1.2, but even for substantially lower aggregate markups the welfare effects of increased product availability remain large and are driven almost entirely by gains from selection.

⁴For example, the model makes clear that changes in the elasticity of substitution and in the shape parameter governing the taste distribution cannot be separately identified using our data on spending. We focus, therefore, on scenarios where only one of these two parameters changes.

able varieties necessary to exactly match empirical trends, and the model consistently delivers substantial unmeasured welfare gains that arise primarily from households choosing product bundles that better suit their unique preferences. Thus, we conclude that while the increase in product availability on its own is not sufficient to fully fit the empirical trends, this increase drives almost all of the welfare effects associated with these trends. The model thus implies a crucial role for product entry in driving the rise of niche consumption, and we provide several additional piece of empirical evidence that support this relationship.⁵

This conclusion that there has been an important welfare relevant increase in product availability arises from fitting trends to a notion of the “average category” in the data. However, we also apply our model separately to each of the the roughly 100 product groups in the data to infer category specific changes in welfare and product availability. We find that the conclusions we arrive at from redoing the analysis category-by-category are broadly similar to those reached when matching average trends. However, there is some interesting heterogeneity across product groups, with the largest gains from selection arising for coffee, disposable diapers, and snacks and negative welfare gains for eggs, cottage cheese, and photographic supplies. Nevertheless, our model implies that the large majority of sectors experienced significant increases in product variety and resulting welfare gains from selection, while changes in other structural parameters again have more limited effects on welfare.

In the final part of the paper, we explore the implications of the rise of niche consumption for market power. Concentration is often used as a proxy for market power, and as described above, markups in our model are endogenous and vary across products so that market power has the potential to move with concentration. So what then happens to aggregate market power in response to the same changes in structural parameters that inferred drove the rise of niche consumption? Surprisingly, not much. In particular, as we demonstrate analytically, increases in product availability do not change the aggregate markup because they have two offsetting effects arising from competition and selection. New products constitute new competition for the incumbents, which causes them to charge lower markups. However, since the new consumed products on average are better tailored to the tastes of households that choose to consume them, they have higher markups than the products they replace. In the aggregate, these two opposing forces exactly offset each other. Changes in the elasticity of substitution and in the shape of the taste distribution have the scope to alter market power, but the changes suggested by our model are not large enough quantitatively to meaningfully change the picture. Our model demonstrates that the significant trends in household and aggregate product concentration need not have led to any changes in aggregate market power.⁶

We proceed as follows. Section 2 discusses the related literature, Section 3 demonstrates the em-

⁵For example, restricting to a balanced panel of products substantially attenuates the rise of niche consumption. And while we again emphasize that measuring variety availability in the data is at best challenging, we find that there is a strong correlation at the category level between empirical measures of observed variety growth and that implied by the model.

⁶Similar conclusions also obtain when we compute aggregate market power trends separately for each of the product groups, although there are a few categories such as photographic supplies where there are non-trivial changes in markups.

empirical divergence between household and aggregate spending concentration, Section 4 develops a theoretical heterogeneous household model to interpret this empirical evidence, and Section 5 concludes.

2 Related Literature

Our work touches on and draws connections between a number of important themes in recent research. Our model in which individual households have CES preferences, heterogeneous Pareto-distributed tastes for different varieties and consume an endogenous subset of these varieties follows [Li \(2018\)](#) and [Arkolakis, Demidova, Klenow, and Rodriguez-Clare \(2008\)](#). We extend this to an environment with heterogeneity in which different households have heterogeneous but potentially correlated tastes for different products. This allows us to speak to the increasing divergence between household and aggregate concentration. Our model implies an important role for expanded product availability in explaining these trends and so our paper is directly related to the vast literature quantifying the effects of variety changes. Our basic approach follows in the tradition of much of the macro and trade literature, building on a CES structure.⁷ However, in our model with heterogeneity and asymmetric preferences, the expansion of product availability benefits consumers through a selection effect above-and-beyond the standard love-of-variety effect which arises in symmetric representative agent models like [Feenstra \(1994\)](#) or [Broda and Weinstein \(2004, 2006\)](#).⁸

Our analysis also relates to a huge literature in industrial organization (IO) quantifying the welfare gains from new varieties.⁹ Within this literature, our result that information on the decomposition of aggregate demand across households can help pin down gains from product availability has close parallels with [Quan and Williams \(2018\)](#). The typical IO approach allows for substantial flexibility in household demand but also typically requires rich data on product characteristics and is computationally intensive to estimate.¹⁰ Our approach requires household level spending data but requires no information on product characteristics, and it delivers simple analytical solutions. This means that unlike the typical IO approach, our methodology scales tractably, and we can apply it to a variety of different sectors and markets. Of course, this tractability comes at the cost of additional parametric structure. While our approach is more flexible than typical symmetric representative agent models and is able to fit spending patterns for most sectors in the data, we view it as a complement rather

⁷For a few examples: [Handbury and Weinstein \(2014\)](#) emphasize the need to account for differences in variety availability when comparing the price level across U.S. cities. [Redding and Weinstein \(2016\)](#) demonstrate that welfare measures in CES environments are biased unless they account for heterogeneity in consumer tastes across products. [Atkin, Faber, and Gonzalez-Navarro \(2018\)](#) use similar scanner data on grocery purchases to calculate the welfare gains associated with entry of global retail chains into the Mexican market.

⁸While representative agent CES models can be rationalized through an underlying discrete choice representation with heterogeneity ([Anderson, de Palma, and Thisse \(1987\)](#)), this requires idiosyncratic tastes to be drawn from an iid Gumbell distribution and aggregates to an environment with symmetry.

⁹cf. [Hausman \(1996\)](#), [Petrin \(2002\)](#), [Brynjolfsson, Hu, and Smith \(2003\)](#)

¹⁰See [Kroft, Laliberté, Leal-Vizcaíno, and Notowidigdo \(2017\)](#) for an alternative sufficient statistic approach to the gains from new varieties.

than a substitute to detailed IO studies of particular markets.

On the empirical side, a recent macro literature has explored the importance of product availability and concentration trends for various empirical phenomenon. In concurrent work, [Michelacci, Paciello, and Pozzi \(2019\)](#) document cyclical fluctuations in household variety adoption and model this phenomenon using a discrete choice model. Their empirical focus is on higher frequency business cycle effects, and their theoretical framework is very different from ours, but they reach similar conclusions about the important role of product selection for welfare. [Argente, Lee, and Moreira \(2018a,b\)](#) show that product introduction plays a key role in understanding patterns of firm growth. [Jaravel \(2019\)](#) argues that innovation and product entry plays an important role in inflation differences across groups. Several important papers document changes in top sales shares and industrial structure, including [Autor, Dorn, Katz, Patterson, and Reenen \(2017\)](#) and [Furman and Orszag \(2015\)](#). Our finding that household and aggregate concentration trends move in opposite directions is reminiscent of [Rossi-Hansberg, Sarte, and Trachter \(2018\)](#), who demonstrate that concentration trends also diverge when comparing measures done at the zip-code and national levels.

Given heterogeneity across households, our structure generates heterogeneous markups because some producers adjust sales by selling more to existing customers while others adjust by selling non-zero amounts to more customers. To our knowledge, [Levin and Yun \(2008\)](#) is the only other paper in the recent literature that emphasizes this mechanism, though it also relates to [Hottman, Redding, and Weinstein \(2016\)](#), who emphasize heterogeneity in the degree to which price declines for one product cannibalize sales for others in multiproduct firms. Our emphasis on differences across firms in the importance of the intensive versus extensive margin contrasts with the more commonly used frameworks for generating variable markups, such as the nested-CES setup in [Atkeson and Burstein \(2008\)](#), linear demand in [Melitz and Ottaviano \(2008\)](#), translog preferences in [Feenstra and Weinstein \(2017\)](#), and [Kimball \(1995\)](#) kinked-demand curves as incorporated in [Gopinath and Itskhoki \(2010\)](#).

Our framework delivers analytical expressions for the full distribution of markups, a topic of increasing focus, such as in the work of [De Loecker and Eeckhout \(2017\)](#), [Edmond, Midrigan, and Xu \(2018\)](#), [Stroebel and Vavra \(2019\)](#) and [Anderson, Rebelo, and Wong \(2018\)](#) and [Bur \(2019\)](#). We note, however, that our model can easily deliver large trends in aggregate and household concentration without requiring any change in aggregate market power. Our work is therefore consistent with the skepticism expressed in [Syverson \(2018\)](#) and [Berry, Gaynor, and Morton \(2019\)](#) of the simple linkage often made between concentration trends and market power.

Finally, although the underlying causes are potentially different, the rise in niche consumption of retail goods parallels the increasing segmentation or polarization witnessed in culture and digital content ([Aguado, Feijoo, and Martinez \(2015\)](#); [Alwin and Tufis \(2015\)](#)), in political ideology ([Pew Research Center \(2014\)](#); [Gentzkow, Shapiro, and Taddy \(2017\)](#)), in jobs and income ([Autor, Katz, and Kearney \(2006\)](#); [Piketty, Saez, and Zucman \(2016\)](#)), and in the geography of where households consume

(Davis, Dingel, Monras, and Morales (2017)).¹¹ Our findings indicate that, along with these other manifestations of fragmentation in modern life, even our grocery purchases increasingly differ from the national average.¹²

3 Diverging Household and Aggregate Concentration

We start this section with a discussion of the aspects of the data that are particularly salient for our analysis, relegating a more detailed description to Appendix A. We then present our key finding that the concentration of household spending across products increased while, at the same time, aggregate concentration among the same goods decreased. Finally, we provide evidence that these trends are associated with product churning.

3.1 AC Nielsen Homescan Data

We use Homescan data from AC Nielsen to measure household-level shopping behavior.¹³ The data set contains a weekly household-level panel for the period 2004-2016. The panel has large coverage, with roughly 170,000 households in over 22,000 zip codes recording prices for almost 700 million unique transactions covering a large fraction of non-service retail spending.

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. Our baseline analysis focuses on annual spending by all households in the Nielsen sample and computes household spending shares across products within product groups, but all results are qualitatively robust to instead calculating household product spending shares within the more disaggregated product modules or within the more aggregated department codes.¹⁴ We focus on the full sample of households for a number of reasons discussed in Appendix D, but this is relatively conservative since the magnitudes of our trends increase when we restrict to a balanced panel of households.¹⁵

In our baseline analysis, we define a product as a Universal Product Code (UPC). Appendix B demonstrates, however, that our results are robust to instead defining a product as a "brand". Nielsen

¹¹The fact that 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 is also consistent with the finding in Bertrand and Kamenica (2018) that cultural distance between rich and poor has not grown over time.

¹²Our findings are also broadly consistent with forecasts of growing importance of "long-tail" consumption (Anderson (2006)) and shows that this phenomenon extends beyond e-commerce to broader retail spending.

¹³These 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. They have also been analyzed recently in Stroebel and Vavra (2019), Jaravel (2019), and Allcott, Diamond, Dube, Handbury, Rahkovsky, and Schnell (2017).

¹⁴All results weight each household using sampling weights provided by Nielsen, which are designed to make the Nielsen panel demographically representative of the broader U.S. population. Appendix Figure A2 shows that aggregate spending growth in our sample tracks government data on aggregate spending growth in comparable categories.

¹⁵In the appendix we also discuss the relevance for our results of additional measurement-related issues, such as the (unimportant) role of online shopping.

assignes UPCs to brands, which are more aggregated than UPCs but are still fairly disaggregated. "Pepsi", for example, is a brand and includes many different flavors and package sizes of the Pepsi drinks. "Caffeine Free - Pepsi", however, is considered a distinct brand. The UPC is our preferred notion of a product in part because UPCs are directly assigned by the manufacturer, whereas the brand variable is constructed by Kilts/Nielsen in a way that involves judgment and may differ across categories and over time. Further, although each generic has a unique UPC, all generics are assigned the same brand in order to preserve the anonymity of the stores in the Nielsen sample. Sales of generics are large and growing, so their inclusion, by construction, distorts concentration measures that define products as brands.¹⁶ Finally, some of our analyses decompose expenditure changes into price and quantity effects, which is straightforward for the case of UPCs but not for brands.

We restrict our analysis to the set of product modules in the data for all years during 2004-2016. We exclude modules that enter or exit since this reflects changes in Nielsen's measurement – not actual household behavior – and could therefore lead to spurious changes in measured concentration. We also exclude fresh produce and other items without barcodes (these are labeled as "magnet" items in the data).

3.2 Household Spending Concentration

We begin our analysis by exploring how the concentration of household spending across products has changed over time. For each household i , UPC j , and product group c we calculate total expenditure $E_{i,j,c,t}$ in year t and associated expenditure share:

$$s_{i,j,c,t} = \left(\frac{E_{i,j,c,t}}{\sum_j E_{i,j,c,t}} \right). \quad (1)$$

Our primary measure of household product concentration for a product category c at time t is the Herfindahl and equals the sum of the square of these expenditure shares:

$$\mathcal{H}_{i,c,t}^{\text{HH}} = \sum_j (s_{i,c,j,t})^2. \quad (2)$$

Next, we take the weighted average across households to generate the Household Herfindahl for product category c :

$$\mathcal{H}_{c,t}^{\text{HH}} = \sum_i \text{share}_{i,c,t} \mathcal{H}_{i,c,t}^{\text{HH}}, \quad (3)$$

where we use weights capturing household i 's share of aggregate spending in category c :

$$\text{share}_{i,c,t} = \frac{\sum_j (\omega_{i,t} E_{i,j,c,t})}{\sum_i \sum_j (\omega_{i,t} E_{i,j,c,t})}, \quad (4)$$

¹⁶See, for example, Dube, Hitsch, and Rossi (2018). Our robustness checks using the brand definition of product exclude generics. Results using UPCs also remain qualitatively robust if we exclude generics.

and where $\omega_{i,t}$ is a household's sampling weight provided to make the Nielsen sample representative of aggregate consumption. Finally, we calculate the overall Household Herfindahl by averaging the category-specific Household Herfindahl in equation (3) across all categories:

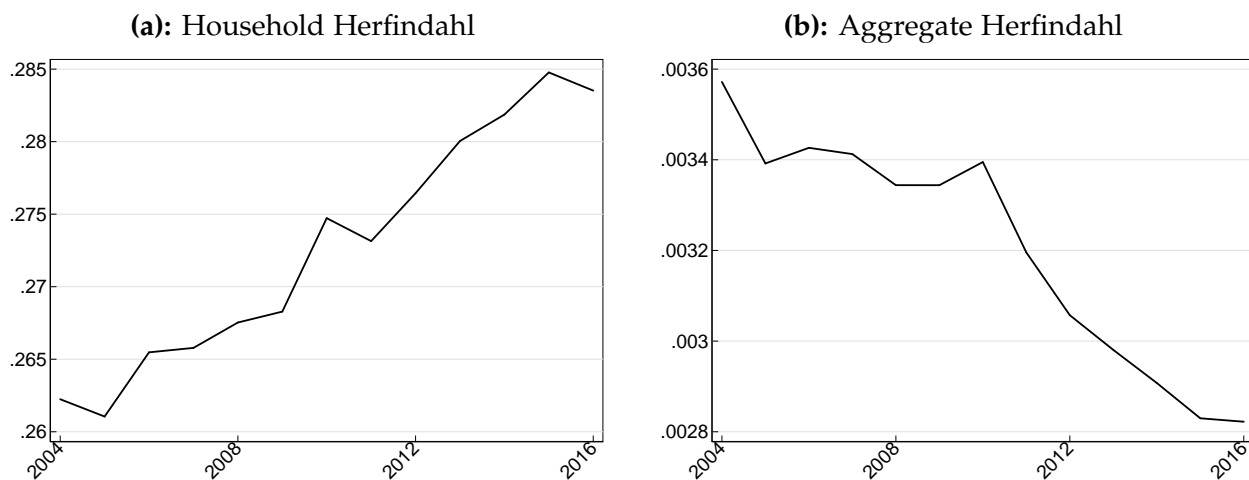
$$\mathcal{H}_t^{\text{HH}} = \sum_c \text{share}_c \mathcal{H}_{c,t}^{\text{HH}}, \quad (5)$$

where share_c is the average share of category c in total spending across our entire sample.

Unlike the weights used in equation (3), we use fixed category spending shares over time in equation (5) to focus on concentration changes occurring within categories, rather than emerging from shifts in spending across categories with different levels of concentration. We do this to better interact with recent interest in changing market power and technological disruption, typically perceived to be occurring within sectors. Our results are qualitatively robust, though a bit noisier, when we allow compositional shifts across categories to influence our concentration measures.

Figure 1a plots $\mathcal{H}_t^{\text{HH}}$ and reveals a nearly monotonic increase in Household concentration from 2004-2016.¹⁷ In Appendix C, we show that this increase in concentration is also associated with a decline in the average number of products consumed per household within a product category. We delay interpreting the quantitative magnitude of these changes until we develop our model in Section 4 but note now that fitting these data with linear trends yields precise and highly significant estimates.

Figure 1: Household Product Concentration



¹⁷All results in the paper hold for alternative concentration measures such as the share of spending accounted for by the top 1 or the top 2 products. We use the Herfindahl as our primary concentration measure as it can be more easily interpreted through the lens of the structural model described in Section 4.

3.3 Aggregate Spending Concentration

What underlies this increase in the concentration of household expenditures? One possible explanation is that there has been an increase in the importance of "super-star products", along the lines of the rise of "super-star firms" documented in [Autor, Dorn, Katz, Patterson, and Reenen \(2017\)](#). This explanation, natural though it may be, finds no support in our data: we demonstrate in this subsection that at the same time the typical household's expenditures have grown more concentrated across products, aggregate spending has in fact become more evenly distributed across these same products.

We sum spending on product j in category c across all households in our data and define the aggregate market share of j in c as:

$$s_{j,c,t} = \frac{\sum_i (\omega_{i,t} E_{i,j,c,t})}{\sum_i \sum_j (\omega_{i,t} E_{i,j,c,t})}, \quad (6)$$

and the Aggregate Herfindahl in category c as:

$$\mathcal{H}_{c,t}^{\text{Agg}} = \sum_j (s_{j,c,t})^2. \quad (7)$$

Just as with the Household Herfindahl, we average these category Herfindahls using fixed category expenditure weights over time to generate the Aggregate Herfindahl of overall spending. Figure 1b plots this Aggregate Herfindahl and shows that the trend in product spending at the aggregate level is the reverse of what we see at the household level: aggregate spending concentration is declining, not rising. How can it be that aggregate concentration is declining if households are individually concentrating their spending on a smaller number of products? These divergent trends imply that households are concentrating more and more spending on their top products over time, but that these top products increasingly differ across households. We view these divergent trends and resulting fragmentation of the product space as characterizing a rise in "niche" consumption.

The decline in aggregate concentration might, at first, seem at odds with the rise in sales concentration measured in Census data by papers including [Autor, Dorn, Katz, Patterson, and Reenen \(2017\)](#). Our aggregate concentration measure, however, captures expenditures at the product level whereas Census-based estimates aggregate products to the producer level.¹⁸ The resulting trends may therefore differ significantly, particularly in the face of changes in the number of goods each manufacturer produces.

In the Appendix, we first show that production concentration measures from the Census for the relevant NAICS categories – "Food Manufacturing" (code 311) and "Beverage and Tobacco Product Manufacturing" (312) – are in fact flat or declining during the years covered in our sample. Next,

¹⁸The categories within which we calculate concentration are also far less aggregated and cover a smaller set of economic activity than what is done in most Census-based studies. Further, our data begin in 2002, far later than the 1970s or 1980s start date commonly found in that literature.

we use a mapping of UPCs to manufacturers to generate a comparable producer-level concentration measure based on the sales in our Nielsen data. We offer a number of important caveats, including that the UPC-to-manufacturer mapping is highly imperfect for this purpose, but nonetheless find similar trends in manufacturer concentration in Nielsen and Census data.¹⁹ We therefore conclude that our results are broadly consistent with the Census-based literature. Whether producer or product concentration is of greater interest depends, of course, on the question at hand. Our theory below will treat each good as produced and marketed independently such that it maps most naturally to our product-based concentration measure.

Finally, it is important to note that while the decline in the number of products consumed by the typical household contributed to the rising Household Herfindahl measure, it is much more difficult to measure the equivalent notion for the aggregate economy. As we show in Appendix C, the existence of thousands of products with tiny amounts of overall sales and incomplete coverage of households and stores in the data render a simple product count highly volatile, dependent on assumptions, and sensitive to measurement error. In contrast, household and aggregate concentration as well as household-level variety statistics are much less sensitive to this issue. We therefore treat the total number of products available for purchase as unobservable, and in Section 4 we show that our model can be used to infer product availability using these other more robust empirical statistics.

3.4 The Pervasive Rise of Niche Consumption

This rise in niche consumption – the increase in the Household Herfindahl and decrease in the Aggregate Herfindahl – is highly robust to various measurement related choices. For example, Appendix Figures A3-A8 show that these divergent trends continue to hold if we exclude generics, compute concentration using more disaggregated categories (modules instead of groups), define products as brands instead of UPCs, use time-varying category weights, use alternative concentration measures instead of the Herfindahl, or focus on a balanced panel of households over time.

Is the rise of niche consumption driven by shifts in the importance of different groups, such as old and young or rich and poor? While there are differences in the level of concentration across different groups, the trends are primarily driven by *within* group variation.²⁰ To show this, we re-calculate annual Household and Aggregate Herfindahls using only expenditures by households with particular demographic characteristics such as income bracket, race, education, and age. Figures 2a and 2b show that rising household and declining aggregate concentration occurs within demographic groups.²¹ The rising Household and falling Aggregate Herfindahls do not simply owe to changes in composition

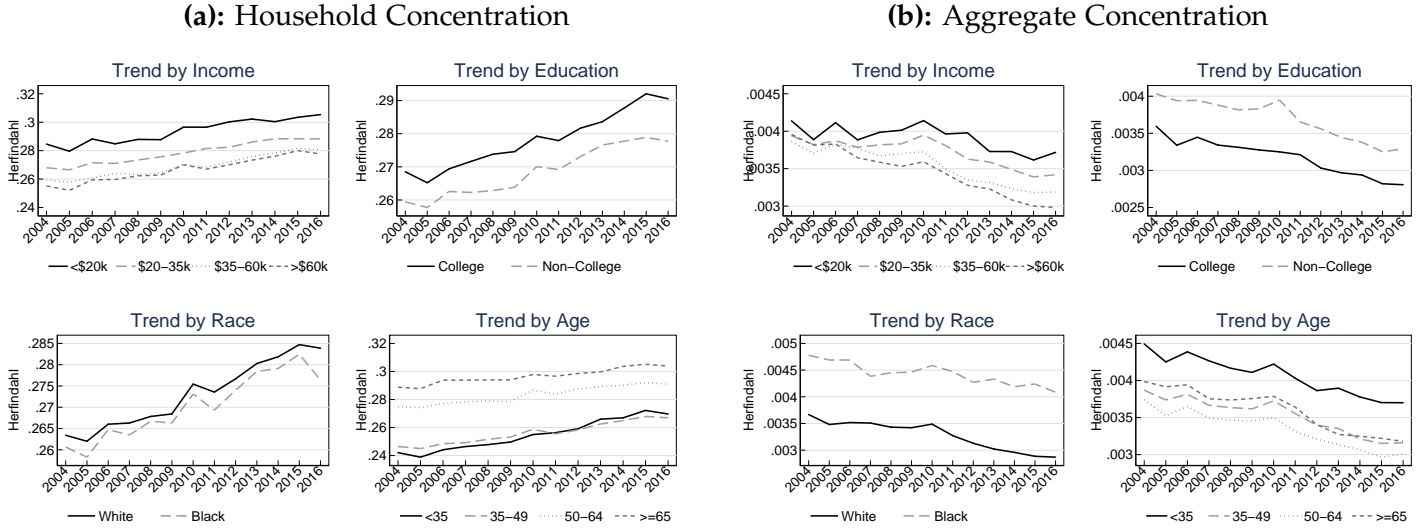
¹⁹We further measure in our data an increase in the share of the top retailers or stores, a finding unrelated to our focus but one that closely resembles the results for the Retail industry in Autor, Dorn, Katz, Patterson, and Reenen (2017).

²⁰While we primarily emphasize trends, the level differences in household concentration are similar to those documented in Hansen and Singh (2015). See also Bornstein (2018) for an analysis of age-specific results.

²¹See also Appendix Figure A9 showing divergent trends within households of different size. We also find diverging trends also obtain when computing concentration in more and less dense cities and cities with higher and lower Republican vote shares as a proxy for political polarization, which is unsurprising in light of Appendix Figure A10 discussed below.

across groups with different levels of concentration.

Figure 2: Trends within demographic group



While Figure 2 focuses on one set of demographics at a time, Appendix Table A1 reports estimated trends in household concentration when we control simultaneously for multiple demographic indicators, including household size. The positive trend in the Household Herfindahl is largely unchanged and remains highly statistically significant.²²

As a simple summary statistic for the prominence of niche consumption, we consider the ratio of the Household Herfindahl to the Aggregate Herfindahl. A higher value for this “niche ratio” means that household consumption is more segmented into different niches. Figure 3 shows that the rise of niche consumption is pervasive across product categories, with three-quarters of product categories exhibiting increases in $\mathcal{H}_{c,t}^{HH}$, eighty percent of product categories exhibiting decreases in $\mathcal{H}_{c,t}^{Agg}$, and growth in the niche ratio in 92 percent of the categories.²³

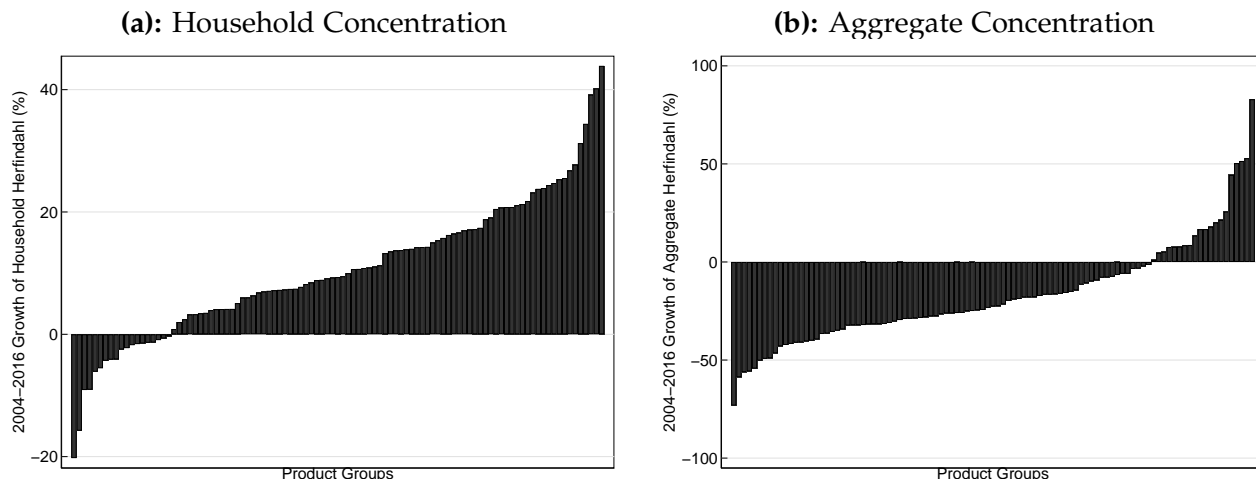
In Appendix Figure A10, we also show that the rise of niche consumption is occurring in the vast majority of locations, implying that shifts in the relative economic importance of cities and regions are not behind our findings. The niche ratio is highest in cities like Chicago, Washington DC, Tampa, Los Angeles, and Boston and lowest in Des Moines, Little Rock, Las Vegas, and “West Texas”, but it is increasing in most locations. Finally, Appendix Figure A11 shows that the rise in niche consumption is found within roughly two-thirds of the individual retailers in our data, so the aggregate patterns we

²²Since \mathcal{H}_i^{HH} is measured for each household i , it is straightforward to regress household concentration measures on a variety of simultaneous demographics which vary across households. A similar exercise for aggregate concentration requires recalculating aggregate market shares and \mathcal{H}^{Agg} separately for each demographic group. This makes these calculations substantially more computationally burdensome. Even more importantly, measurement error in aggregate market shares increases rapidly for more narrow demographic groups.

²³To improve visual exposition, Figures 3 and 4a drop 5 outlier categories whose variety counts more than double or decrease by more than 50 percent from 2004-2016: “Frozen Juices”, “Yeast”, “Canning Supplies”, “Greeting Cards” and “Photographic Supplies”. This does not affect any conclusions.

observe are not simply driven by shifts in where households shop.²⁴ Together these results all imply that whatever forces are driving the rise in niche consumption, they are pervasive across demographics, geographies, retail chains, and product categories.

Figure 3: 2004-2016 Concentration Growth Within Category



The level of the niche ratio is highest in “Cosmetics” and “Fragrances-Women” and is lowest for “Charcoal” and “Dough Products”. The rise in niche consumption is pervasive, but it is also clear from Figure 3 that there is substantial heterogeneity across categories in the extent of its ascent. The niche ratio has grown most rapidly for “Coffee”, “Hardware, Tools”, “Fresheners and Deodorizers”, and “Disposable Diapers”. It has declined by most for “Cottage Cheese”, “Eggs”, “Milk”, and “Bread and Baked Goods”.

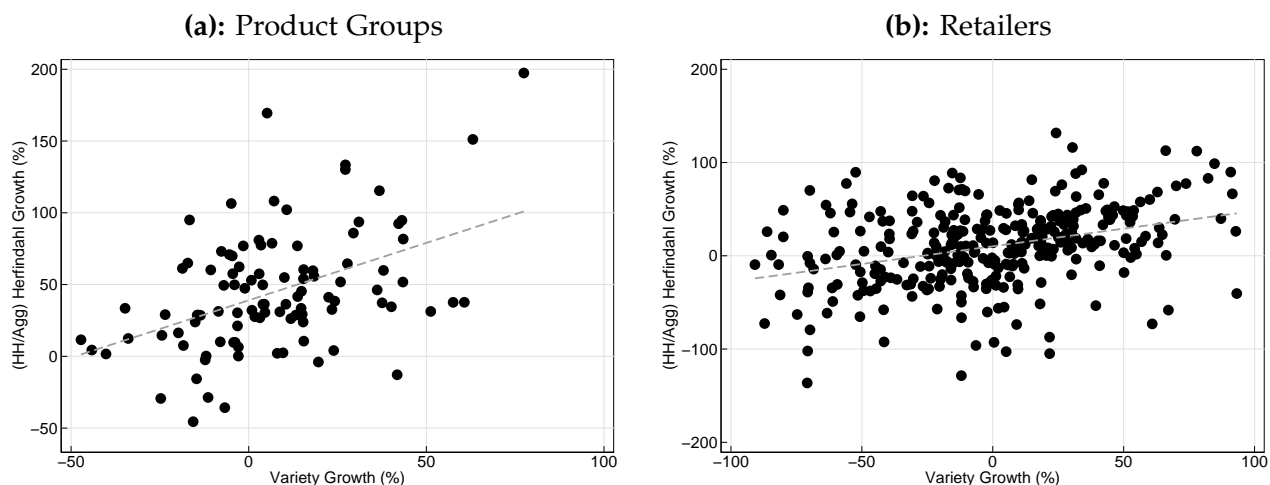
3.5 The Role of Product Churn

Interestingly, there is a common observable linking together the categories with the most rapid increases in niche ratios: they are also the categories with the fastest growth in the number of aggregate products, measured using the same assumptions as were used for Figure ?? above. We emphasize this relationship as it will be central to the mechanism in our model in Section 4 and its implications for welfare. In particular, Figure 4a shows that categories with 50 percentage points more growth in the total number of products sold had, on average, 40 percentage points more growth in their niche ratios, with the relationship statistically significant at the 1 percent level. Figure 4b shows that a similar relationship also holds when comparing across retailers: retailers with 50 percentage points more growth in the number of products sold exhibited roughly 20 percentage points more growth in their niche

²⁴We also show that our results are not driven by trends in the frequency of shopping and bulk purchasing (Coibion, Gorodnichenko, and Koustas (2017)).

ratios, with the relationship again significant at the 1 percent level.²⁵ The relationship between variety growth and the niche ratio becomes even steeper if we weight retailers by size.

Figure 4: Growth in Number of Products vs. Growth in Niche Ratio



We now provide additional evidence that product churn plays a key role in the rise of niche consumption by comparing concentration trends measured only among “continuing” products that are purchased by a household in two consecutive years with those measured using all spending by that household.²⁶ For each household i that is observed in both t and $t + 1$, we measure concentration of “continuing products” by using only those that are purchased by that household in both t and $t + 1$. These continuing products account for about 30 percent of transactions and 40 percent of spending. We also calculate Herfindahls for those same households using all their spending. We form an index by chaining together changes in these Herfindahls from t and $t + 1$ and pin down the level using the values in the initial period.²⁷ Figure 5a shows the upward trend in household concentration is much stronger when using all UPCs than when restricting to continuing products, growing by 29 percent compared to 5 percent.²⁸ This implies a large role for product entry and exit in generating household concentration increases. Figure 5b shows that when focusing only on continuing products, aggregate concentration actually rises instead of declines.²⁹

²⁵To reduce the influence of outliers, we exclude retailers with absolute log variety changes above 2, which drops 6 out of 334 retailers. Results are similar for alternative thresholds. The panel of retailers is unbalanced, and growth rates for the remaining 328 retailers are calculated from their first to their last observation in the sample. Results are very similar if we instead calculate growth rates from 2004-2016 for the 179 retailers which are in the sample continuously.

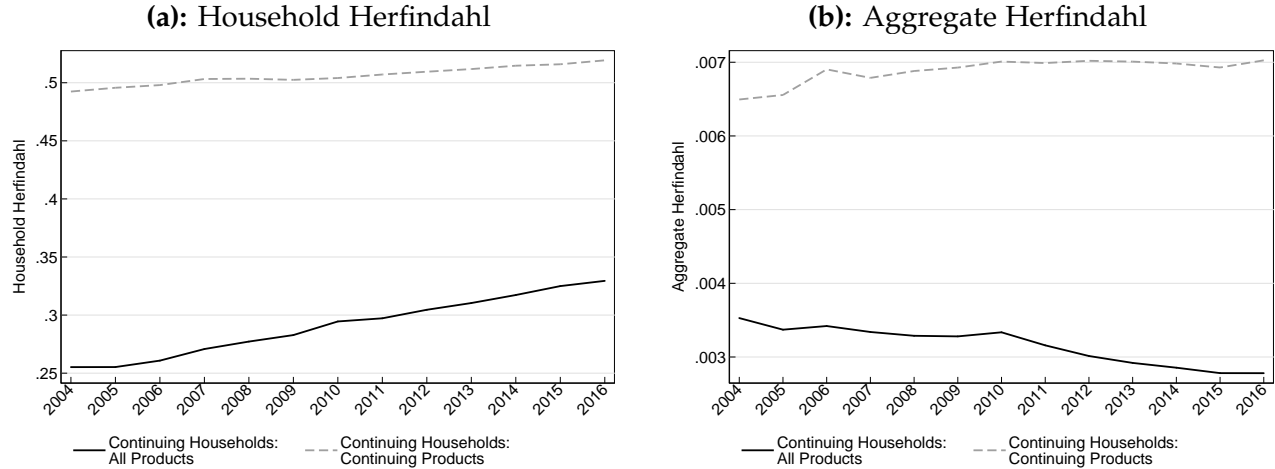
²⁶For these results, therefore, we move to a panel specification since we require at least two time periods for a household to measure which products for that household are new and which are continuing.

²⁷We report cumulative changes rather than levels since this procedure results in two levels of the Herfindahl for each year (one initial and one continuation year), so the levels are not meaningful.

²⁸The trend for “All Products” is larger than that in Figure 1a as Figure 5a is calculated using within-household variation. See Appendix D for details.

²⁹While this analysis shows that the diverging concentration trends are largely driven by extensive rather than intensive margin effects, for continuing products we can nevertheless decompose intensive margin concentration trends into P vs. Q effects. Appendix Figure A12 shows that changes in quantity rather than price are more important.

Figure 5: 2004-2016 Concentration growth for continuing vs. all products



4 Modeling the Rise of Niche Consumption

In this section, we develop a model able to match the rise of niche consumption documented in Section 3 and we use it to identify key driving forces and understand the resulting implications for welfare and markups. Many standard models are not useful for studying a simultaneous increase in household concentration and decline in aggregate concentration because they either assume that all households consume a single product, assume that household tastes are symmetric across products, or assume that all households are identical. In our model, households choose how many products to consume, spend different amounts on each good, and differ from other households in their choice of which products to buy.

Following Li (2018), we assume that households must pay a fixed cost per consumed product, which implies they only consume a subset of available products despite their CES preferences that embed a “love of variety”. As in Arkolakis, Demidova, Klenow, and Rodriguez-Clare (2008), we assume each household’s tastes for products, adjusted for price, are distributed Pareto, which allows us to write the Household Herfindahl analytically. Further, we introduce a rank function that implies the preference ordering of products will differ across households, which allows us to write the Aggregate Herfindahl analytically.

Using the analytical expressions for the Household and Aggregate Herfindahl, we confront the model with empirical trends from 2004-2016 and back out the implied driving forces. We find that an increase in the number of available products is required to quantitatively match the rise of niche consumption. In the model, this increase leads to significant welfare gains as it implies that consumers can choose a consumption bundle better tailored to their tastes without raising their fixed cost expenditures.

Finally, the model features heterogeneous markups across products. Growth in the sales of products with larger aggregate market shares primarily reflect growth in spending by existing customers, or intensive margin adjustment. By contrast, growth in the sales of small products are more likely to come from the addition of new customers, or extensive margin adjustment. Intensive and extensive margin adjustments are characterized by different elasticities of demand, and this results in heterogeneous markups. While changes in the number of products shift markups across products, we demonstrate that they do not impact the aggregate degree of market power in the economy.

4.1 Household Problem

We assume that a continuum of households $i \in [0, 1]$ spend E on a continuum of varieties $k \in [0, N]$ to maximize:³⁰

$$U_i = \left(\int_{k \in \Omega_i} (\gamma_{i,k} C_{i,k})^{\frac{\sigma-1}{\sigma}} dk \right)^{\frac{\sigma}{\sigma-1}} - F \times (|\Omega_i|)^\epsilon, \quad (8)$$

where Ω_i is the set of products consumed by i (with $|\Omega_i| \leq N$), $\gamma_{i,k}$ is a household-specific taste for product k , and the term multiplied by F captures a fixed cost that increases exponentially in the measure of varieties consumed.

We write the price of product k as p_k , so $\tilde{\gamma}_{i,k} = \gamma_{i,k}/p_k$ captures the price-adjusted taste of household i for k . We assume price-adjusted tastes are distributed Pareto:

$$Pr(\tilde{\gamma}_{i,k} < y) = G(y) = 1 - (y/b)^{-\theta},$$

where $y \geq b > 0$ and where we assume $\theta > 2(\sigma - 1)$. Since larger θ means a flatter distribution of tastes, the latter condition simply ensures that tastes are not "too concentrated" relative to σ and that the model delivers a finite Household Herfindahl. We also assume $\epsilon > 1/(\sigma - 1) - 1/\theta$, which implies that higher fixed costs F lead to less purchased products $|\Omega|$. Household i will consume the set of goods with $\tilde{\gamma}_{i,k} \in [\tilde{\gamma}^*, \infty)$ for some $\tilde{\gamma}^* \geq b$.

The ideal price index in this environment will be equal for all households and is defined as:

$$\begin{aligned} P_i = P &= \left(\int_{k \in \Omega_i} (\tilde{\gamma}_{i,k})^{\sigma-1} dk \right)^{\frac{1}{1-\sigma}} \\ &= \underbrace{\left(1 + \frac{1-\sigma}{\theta} \right)^{\frac{1}{\sigma-1}} b^{-1}}_{\text{Average Price}} \times \underbrace{(|\Omega_i|)^{\frac{1}{1-\sigma}}}_{\text{Variety Gains}} \times \underbrace{\left(\frac{|\Omega_i|}{N} \right)^{\frac{1}{\theta}}}_{\text{Selection Effects}}. \end{aligned} \quad (9)$$

The price index has three terms, each with an intuitive interpretation. We refer to the first term as the average price since it summarizes the full distribution of price-adjusted tastes for available products

³⁰To ease notation, we do not index k by i , but importantly note that the same k may represent a different actual product for each different households. This is unimportant for the analysis of individual households, but will be crucial when we move to the aggregate analysis.

as if there were a single purchase price for one unit of the full bundle. It varies with the shape θ and scale b of the Pareto distribution as well as with the elasticity of substitution σ . The second term is the standard love-for-varieties term in CES models, which decreases with the measure of consumed products and with the elasticity of substitution (given $|\Omega_i| > 1$). Finally, the third term represents a selection effect from the fact that when households only consume a subset Ω_i of the full measure N of products, they choose the subset they like best. This term decreases in the share of products that are consumed and in the extent to which households prefer some products to others.

The index reduces to more standard expressions in special cases. For example, consider $\theta \rightarrow \infty$, which implies that households value all products identically at b , i.e. $\tilde{\gamma}_{i,k} = b$ for all i and k . In such a case, the expression reduces to $b^{-1}|\Omega_i|^{1/(1-\sigma)}$, which is the standard price index for symmetric CES preferences. Alternatively, imagine some products are preferred to others, $\theta < \infty$, but all products are nonetheless purchased, $\Omega_i = N$. In this case, the last term reduces to 1 as there are no selection effects and the average price term fully captures impact of heterogeneity in the desirability of the products.

The properties of the CES price index imply we can re-write equation (8) as:

$$U_i = \frac{E}{P} - F \times (|\Omega_i|)^\epsilon.$$

Consumers choose $|\Omega_i|$ to maximize utility. The first order condition implies that the optimal number of products is:

$$|\Omega_i| = |\Omega| = \left(\frac{\left(\frac{1}{\sigma-1} - \frac{1}{\theta}\right) \left(1 + \frac{1-\sigma}{\theta}\right)^{\frac{1}{1-\sigma}} N^{\frac{1}{\theta}}}{\tilde{F}\epsilon} \right)^{\left(\epsilon - \frac{1}{1-\sigma} + \frac{1}{\theta}\right)^{-1}}, \quad (10)$$

where $\tilde{F} = F/(bE)$ is a parameter which shifts spending, aggregate prices, and variety costs.³¹ Importantly, the optimal choice of varieties yields a "cutoff" taste $\tilde{\gamma}^*$ that satisfies: $G(\tilde{\gamma}^*) = 1 - \frac{|\Omega|}{N}$, and the share of household i 's expenditure on variety k is then given by

$$s_{i,k} = \begin{cases} (P\tilde{\gamma}_{i,k})^{\sigma-1}, & \tilde{\gamma}_{i,k} > \tilde{\gamma}^* \\ 0, & \tilde{\gamma}_{i,k} \leq \tilde{\gamma}^* \end{cases} \quad (11)$$

with $\int_k s_{i,k} dk = 1$.

4.2 Household Herfindahls

Given equation (11), it follows that the Household Herfindahl \mathcal{H}^{HH} will be equal for all i and can be written as:

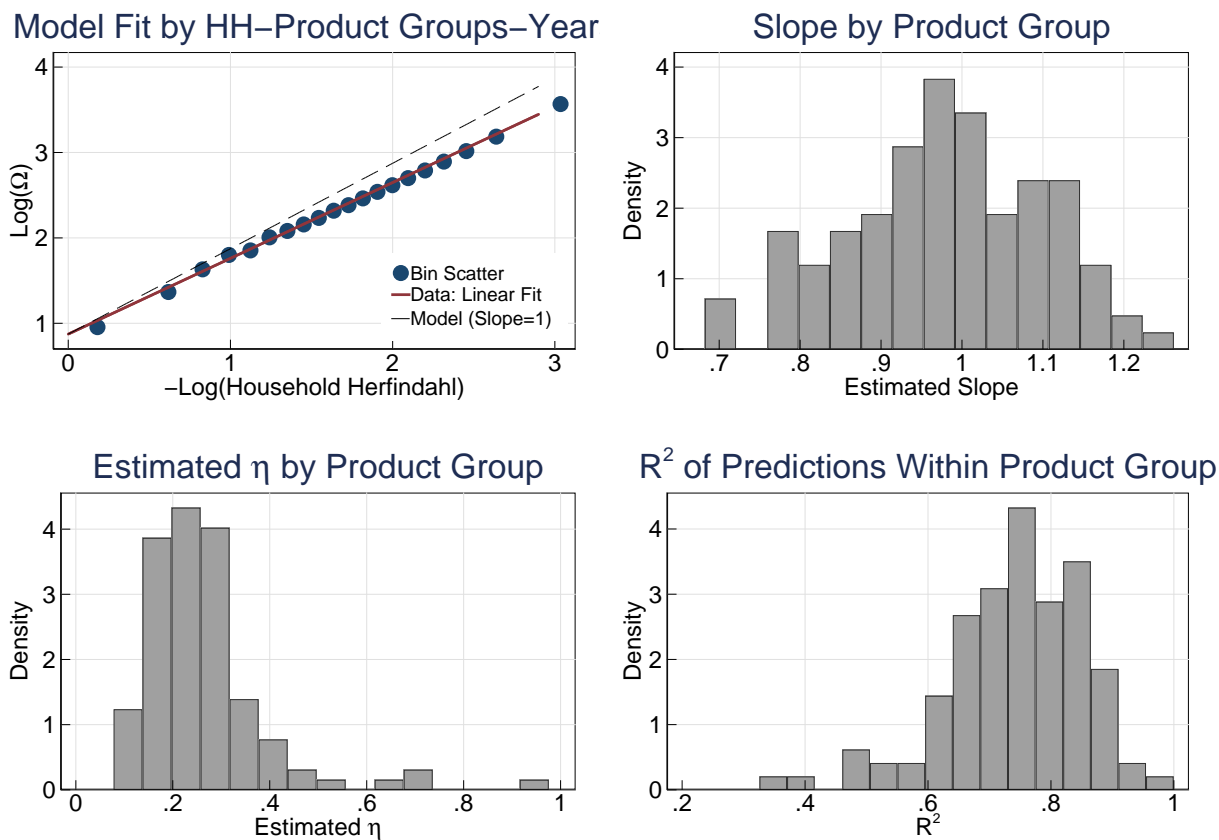
$$\mathcal{H}^{\text{HH}} = \mathcal{H}_i^{\text{HH}} = \int_{k \in \Omega_i} (s_{i,k})^2 dk = N \int_{\tilde{\gamma}_i^*}^{\infty} (P_i \tilde{\gamma}_{i,k})^{2(\sigma-1)} dG(y)$$

³¹When $N = 1$, this expression is the same as that in Li (2018) after substituting in his special case for b .

$$= \frac{(\eta + 1)^2}{4\eta} \frac{1}{|\Omega|'} \quad (12)$$

where we introduce the variable $\eta = 1 - 2(\sigma - 1)/\theta$. The above parameter restrictions imply $\eta \in (0, 1)$. For fixed θ and σ , which implies fixed η , household concentration declines monotonically with the number of consumed varieties. And for fixed $|\Omega|$, concentration declines monotonically with η . All else equal, flatter taste distributions (higher θ) or less substitutability across products in preferences (lower σ) reduce Household Herfindahls.

Figure 6: Model Fit on Household-Category Data



How well does this model fit household spending data? Interpreting our model as applying to each household's spending decisions for a given product category c in a given year, we have the testable prediction that $\mathcal{H}_{i,c}^{HH}$ is proportional to $1/|\Omega_{i,c}|$. Indeed, when we pool categories, years, and households and regress $\ln |\Omega_{i,c}|$ on $-\ln \mathcal{H}_{i,c}^{HH}$ (with category-year fixed effects), we get a coefficient of 0.89 – close to the model-consistent value of 1 – and a large R^2 of 0.82. The upper left panel of Figure 6 shows a binscatter (with category-year fixed effects) of the 54 million observations underlying this regression to demonstrate that linearity with a coefficient of 1 is a close approximation to the raw data.³² In the upper right panel, we estimate these regressions separately for each category in 2016 and

³²This specification has large explanatory power even though it only allows η to vary across category-years and not across

plot a histogram of the estimated slopes. The values are largely clustered around the model-consistent value of 1.

Next, rather than estimating the slope, we constrain it to equal 1 and back out the η values implied for each category. The model imposes the restriction that $0 < \eta < 1$ and the bottom left panel of Figure 6 shows that this restriction is satisfied in every category. The values of η range from lows of 0.08 (Baby Food) and 0.10 (Carbonated Beverages) to highs of 0.69 (Greeting Cards) and 0.97 (Yeast).³³ Finally the lower right panel shows that the R^2 's from these restricted regressions are generally high. Overall, we conclude that the empirical relationship between household-level concentration measures and the number of consumed products is consistent with the relationships implied in our model.

4.3 Aggregation

In order to account for divergent trends in household and aggregate concentration measures, we must specify how tastes for particular products differ across households. We index all products in the economy with $j \in [0, N]$, and assume each household assigns each product a “rank”, where lower ranks indicate higher price-adjusted tastes. Households will consume all goods which they rank less than or equal to $|\Omega|$.

We introduce the following rank function for each household i :

$$r_{i,j} = (1 - \alpha)j + \alpha x_{i,j}, \quad (13)$$

where j identifies a common aggregate rank for a product, $x_{i,j}$ is an i.i.d. draw from the uniform distribution with support $[0, N]$ representing a household-specific taste component, and $\alpha \in (0, 1)$. If α is close to zero, the model approximates a representative agent model where all households rank products in the same order. If α approaches 1, tastes are purely idiosyncratic and resulting aggregate spending will be evenly distributed over all consumed products even if individual household tastes are very concentrated. Thus, even though all households have identical *distributions* of taste-adjusted prices, this rank function allows for different households to have different ranks for the exact same product j .

To compute the aggregate spending share on product j , we need to know the cumulative distribution function (CDF) of product ranks $R(r)$, integrating over all household and products. Without loss

households. With arbitrary heterogeneity in η across households within category-years, there would be as many parameters as observations so it would be trivial to perfectly fit the data.

³³The value of 0.08 for baby food implies that the typical household in this category has spending which is almost 4 times more concentrated than if that household spent evenly on all the baby food products they consumed, while the value of 0.97 for yeast implies that household spending in that category is essentially evenly divided across products. With homogeneous tastes across products (i.e. $\theta \rightarrow \infty$) – the setup in many standard models – we cannot capture this large extent of sectoral heterogeneity in concentration as $\eta \rightarrow 1$.

of generality, we assume $\alpha < 1/2$ and write:³⁴

$$R(r) = \begin{cases} \frac{1}{2} \left(\frac{r}{N}\right)^2 \frac{1}{\alpha(1-\alpha)}, & 0 \leq r < \alpha N \\ \frac{r}{N} \frac{1}{1-\alpha} - \frac{1}{2} \frac{\alpha}{1-\alpha}, & \alpha N \leq r < (1-\alpha)N \\ -\frac{1}{2} \left(\frac{r}{N}\right)^2 \frac{1}{\alpha(1-\alpha)} + \frac{r}{N} \frac{1}{\alpha(1-\alpha)} - \frac{1}{2} \left(\frac{\alpha}{1-\alpha} + \frac{1-\alpha}{\alpha}\right), & (1-\alpha)N \leq r \leq N. \end{cases} \quad (14)$$

Note that this CDF satisfies the properties that $R(0) = 0$, $R(N) = 1$, is continuous at $r = \alpha N$ and $r = (1-\alpha)N$, and is monotonically increasing.

There are three distinct regions in $R(r)$ with different functional forms. If households only consume goods with ranks in the first region, this implies that there is no single product in the economy that is purchased by all households. If households consume so many varieties that some have ranks in the second region, this implies that at least one product is purchased by everyone. Finally, if even the worst possible product in the economy is purchased by at least one household, then the ranks of some consumed goods will fall into the third region.³⁵ As long as $0 \leq \frac{|\Omega|}{N} < \frac{\alpha}{2(1-\alpha)} < \frac{1}{2}$, it can be shown that all consumed products in the economy will have an r value confined to the first region of $R(r)$. This is the empirically relevant region of the parameter space, since the number of varieties purchased by an individual household is orders of magnitude less than the aggregate number of varieties, and there are no varieties in the data that are consumed by all households. To simplify analytical solutions, we thus impose this parameter restriction for the remainder of the analysis.

Noting that $\tilde{\gamma}_{i,j} = G^{-1}(1 - R(r_{i,j}))$, the spending share of household i that is dedicated to product j can be written as a function of j 's rank:

$$s_{i,j} = P^{\sigma-1} \tilde{\gamma}_{i,j}^{\sigma-1} = (Pb)^{\sigma-1} (R(r_{i,j}))^{-\frac{\sigma-1}{\theta}} = \frac{\eta+1}{2} N^{\frac{\eta-1}{2}} |\Omega|^{-\frac{\eta+1}{2}} (R(r_{i,j}))^{\frac{\eta-1}{2}}, \quad (15)$$

if $R(r_{i,j}) \leq |\Omega|/N$ and zero otherwise. To determine the products for which the share $s_{i,j}$ in equation (15) jump from positive to zero, we solve for the rank of the marginal, or least-preferred, variety that is consumed in positive quantities by household i . Note that this good's identity will differ across households, but its rank r^* will be the same and satisfies $R(r^*) = |\Omega|/N$. Substituting into equation (14) under the assumption that $0 \leq \frac{|\Omega|}{N} < \frac{\alpha}{2(1-\alpha)}$, we get:

$$r^* = (2\alpha(1-\alpha)|\Omega|N)^{\frac{1}{2}}. \quad (16)$$

³⁴Replacing α with $1-\alpha$ in all instances in equation (14) yields the corresponding $R(r)$ for the alternative case of $\alpha > 1/2$. Furthermore, this leaves the rank function unchanged for the first of the three regions of $R(r)$, which will be the focus of our analysis.

³⁵More specifically, the product with the best aggregate taste shock is $j = 0$. The worst possible idiosyncratic rank for this product occurs when $x_{i,j} = N$, in which case $r = \alpha N$, so if we are in the first region of the parameter space, even the best product is not purchased by some households. Conversely, the product with the worst aggregate taste shock is $j = N$. The best possible idiosyncratic rank for this product occurs when $x_{i,j} = 0$, in which case $r = (1-\alpha)N$. This means that if we are in the third region of the CDF, this worst product will still be consumed by some household.

Under our parameter restrictions, individual households each consume only a fraction of the total products available N , but the exact products consumed will differ across households. However, even when aggregating across all households, there are some products which are consumed by no households. This means that for the economy as a whole, there is a difference between the measure of available goods N and the measure of goods that are actually consumed, which we denote j^* . This marginal consumed good for the economy as a whole, j^* , is that j for which the best possible idiosyncratic taste draw – a draw of zero – yields rank r^* for the household with that zero draw. Solving for this cutoff, $j^* = r^* / (1 - \alpha)$, we get:

$$j^* = \left(\frac{2\alpha|\Omega|N}{1-\alpha} \right)^{\frac{1}{2}}. \quad (17)$$

Importantly, since $r_{i,j}$ is strictly increasing in j , all goods with $j < j^*$ will have positive aggregate sales and all goods with $j \geq j^*$ will have zero aggregate sales. Finally, substituting in the definition of the rank function from equation (13) into the expression (16), and using the definition of j^* in equation (17), we can write the highest value or “cutoff” random draw x_j^* that yields positive consumption of j as:

$$x_j^* = \frac{1-\alpha}{\alpha} (j^* - j). \quad (18)$$

4.4 The Aggregate Herfindahl

We now integrate spending shares across households i to get the aggregate spending share on good j :

$$\begin{aligned} s_j &= \frac{1}{\int_i E d_i} \int_i E s_{i,j} d_i \\ &= \frac{\eta+1}{2} N^{\frac{\eta-1}{2}} |\Omega|^{-\frac{\eta+1}{2}} \int_0^{\frac{1-\alpha}{\alpha}(j^*-j)} 2^{\frac{1-\eta}{2}} N^{1-\eta} \left(\alpha (1-\alpha)^{\frac{1-\eta}{2}} \right) ((1-\alpha)j + \alpha x_{i,j})^{\eta-1} \frac{dx}{N} \\ &= (\eta+1) (2N|\Omega|)^{-\frac{\eta+1}{2}} (\alpha(1-\alpha))^{\frac{1-\eta}{2}} \int_0^{\frac{1-\alpha}{\alpha}(j^*-j)} ((1-\alpha)j + \alpha x_{z,i})^{\eta-1} dx \\ &= \frac{\eta+1}{\eta} \left(\frac{2\alpha N|\Omega|}{1-\alpha} \right)^{-\frac{\eta+1}{2}} ((j^*)^\eta - j^\eta) \\ &= \frac{\eta+1}{\eta j^*} \left(1 - \left(\frac{j}{j^*} \right)^\eta \right). \end{aligned} \quad (19)$$

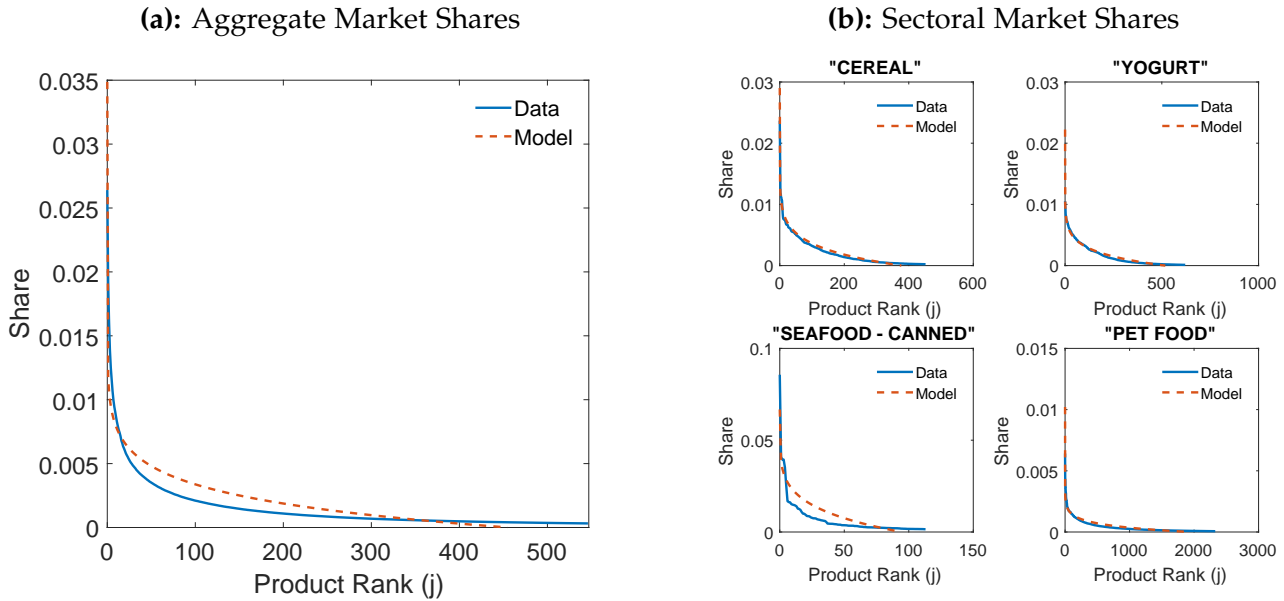
Using equation (19), we immediately obtain the Aggregate Herfindahl:

$$\begin{aligned} \mathcal{H}^{\text{Agg}} &= \int_{j=0}^{j^*} s_j^2 dj = \left(\frac{\eta+1}{\eta j^*} \right)^2 \int_{j=0}^{j^*} \left(1 - \left(\frac{j}{j^*} \right)^\eta \right)^2 dj \\ &= \left(\frac{\eta+1}{\eta j^*} \right)^2 j^* \left(1 - \frac{2}{\eta+1} + \frac{1}{2\eta+1} \right) \\ &= \frac{2(\eta+1)}{(2\eta+1)} \frac{1}{j^*} \end{aligned}$$

$$= \frac{2(\eta + 1)}{(2\eta + 1)} \left(\frac{1}{2\tilde{N}|\Omega|} \right)^{\frac{1}{2}}, \quad (20)$$

where we define $\tilde{N} = N\alpha/(1 - \alpha)$. Aggregate concentration declines monotonically with \tilde{N} . For fixed θ and σ , aggregate concentration declines monotonically with the number of consumed products. And for fixed $|\Omega|$, concentration declines monotonically with η . Importantly, changes in $|\Omega|$ and η move the Household Herfindahl and Aggregate Herfindahl in the same direction. As we discuss in the next subsection, this imposes strong restrictions on the set of possible forces which can explain the opposite empirical trends for \mathcal{H}^{HH} and \mathcal{H}^{Agg} and implies an important role for increases in \tilde{N} .

Figure 7: Model Fit on Aggregate Category Data



How well do these model-based relationships fit aggregate sales distributions in the data? To assess this, we start by measuring $|\Omega|$ directly in the data and then solve for the two remaining free parameters, η and \tilde{N} , to match \mathcal{H}^{HH} and \mathcal{H}^{Agg} in equations (12) and (20). Figure 9 then plots the market share distribution across products implied by our model in equation (19) (the red dashed line) against the actual market share distribution in the data (the solid blue line). We do this for total spending in Figure 7a as well as separately for a number of product categories in Figure 7b. In several categories such as cereal and yogurt, the model fits extremely well, while it is notably less successful in others such as greeting cards or canned seafood. Overall, however, we consider these good fits as validating our use of the model, particularly given the distributions are fully determined by only three parameters and reflect parametric assumptions and functional forms chosen largely for analytical convenience.

4.5 Elasticities of Demand, Markups, and Aggregate Profits

The previous sections develop analytical expressions for our key empirical objects: \mathcal{H}^{HH} , \mathcal{H}^{Agg} and $|\Omega|$. We will show in Section 4.6 that these expressions can be used to draw important conclusions about the forces driving the rise of niche consumption, but we defer this analysis. This is because the calibration of our model and the resulting welfare implications of the rise in niche consumption will depend on the elasticity of substitution. In typical CES environments, the elasticity of demand and markups are fully determined by the exogenous elasticity of substitution σ . By contrast, we show in this section that the elasticity of demand in our model depends both on this standard “intensive margin” force as well as on an endogenous “extensive margin” force that arises from the possibility for products to gain new customers (or lose existing ones). Since these forces are of different importance for products with different aggregate market shares, the model generates heterogeneous elasticities and resulting markups. This also implies that existing estimates of the elasticity of substitution (e.g. [Broda and Weinstein \(2004\)](#)) cannot be applied in our context.

To solve for the price elasticity of aggregate demand for product j , we start by expressing its total sales as the integral of each household’s spending on j , taken over all households:

$$s_j = \frac{1}{N} \int_0^{x_j^*} s_{i_x,j} dx, \quad (21)$$

where we use the notation $s_{i_x,j}$ to denote the spending share of a household with taste draw on product j equal to x . Since j will only be purchased by those households with a sufficiently high idiosyncratic taste for it, we need only integrate from households drawing $x_{i,j} = 0$ to the marginal household that draws $x_{i,j} = x_j^*$.

We take the partial derivative of s_j in equation (21) with respect to p_j to get:

$$\frac{\partial s_j}{\partial p_j} = \frac{1}{N} \left(\int_0^{x_j^*} \frac{\partial s_{i_x,j}}{\partial p_j} dx + s_{i_{x_j^*},j} \frac{\partial x_j^*}{\partial p_j} \right), \quad (22)$$

where the right hand side of equation (22) follows from Leibniz’s rule. The first term can be solved using equation (15) as:

$$\frac{\partial s_{i_x,j}}{\partial p_j} = \frac{\partial P^{\sigma-1} p_j^{1-\sigma} \gamma_{i,j}^{\sigma-1}}{\partial p_j} = (1-\sigma) \frac{s_{i_x,j}}{p_j}, \quad (23)$$

where we take the aggregate price index P as fixed. Moving on to the second term, we can evaluate equation (15) at the marginal household with taste x_j^* to get:

$$s_{i_{x_j^*},j} \frac{\partial x_j^*}{\partial p_j} = \frac{\eta+1}{2} N^{\frac{\eta-1}{2}} |\Omega|^{-\frac{\eta+1}{2}} (R(r^*))^{\frac{\eta-1}{2}} \frac{\partial x_j^*}{\partial p_j}. \quad (24)$$

Substituting equations (23) and (24) back into equation (22), we get:

$$\begin{aligned}
\frac{\partial s_j}{\partial p_j} &= (1 - \sigma) \frac{1}{p_j} \frac{1}{N} \int_0^{x_j^*} s_{i,x,j} dx + \frac{1}{N} s_{i,x^*,j} \frac{\partial x_j^*}{\partial p_j} \\
&= (1 - \sigma) \frac{s_j}{p_j} + \frac{1}{N} \frac{\eta + 1}{2} N^{\frac{\eta-1}{2}} |\Omega|^{-\frac{\eta+1}{2}} (R(r^*))^{\frac{\eta-1}{2}} \frac{\partial x_j^*}{\partial p_j} \\
&= (1 - \sigma) \frac{s_j}{p_j} + \frac{1}{N} \frac{\eta + 1}{2} N^{\frac{\eta-1}{2}} |\Omega|^{-\frac{\eta+1}{2}} \left(\frac{1}{2} N^{-2} \alpha^{-1} (1 - \alpha) (j^*)^2 \right)^{\frac{\eta-1}{2}} \frac{\partial x_j^*}{\partial p_j} \\
&= (1 - \sigma) \frac{s_j}{p_j} + \frac{\eta + 1}{2N|\Omega|} \frac{\partial x_j^*}{\partial p_j}.
\end{aligned} \tag{25}$$

We cannot directly consider a price change, required to exactly solve for $\partial x_j^*/\partial p_j$, because the full distribution of price-adjusted tastes in our model is set exogenously and must be maintained as Pareto. Instead, to *approximate* the change in mass of households consuming a product when that product's price increases, we start with the relationship:³⁶

$$\frac{1}{2} \left(\frac{r_{i,j}}{N} \right)^2 \frac{1}{\alpha(1-\alpha)} = R((1-\alpha)j + \alpha x_{i,j}) = 1 - G\left(\frac{\gamma_{i,j}}{p_j}\right) = b^\theta \gamma_{i,j}^{-\theta} p_j^\theta, \tag{26}$$

and differentiate to yield:

$$\frac{r_{i,j}}{1-\alpha} \frac{1}{N^2} \frac{\partial x_{i,j}}{\partial p_j} = \theta b^\theta \gamma_{i,j}^{-\theta} p_j^{\theta-1}, \tag{27}$$

where we've substituted $\partial r_{i,j}/\partial x_{i,j} = \alpha$. We then evaluate equation (27) at $r_{i,j} = r^*$ and $\tilde{\gamma}_{i,j} = \tilde{\gamma}^*$ using equation (16), and add a minus sign to reflect the fact that increase in the price of good j should reduce the set of households purchasing that good, to get:

$$\frac{dx_j^*}{dp_j} = -\frac{\theta}{j^*} |\Omega| N \frac{1}{p_j}. \tag{28}$$

Inserting this into equation (25), we have:

$$\begin{aligned}
\frac{\partial s_j}{\partial p_j} &= (1 - \sigma) \frac{s_j}{p_j} - \frac{\eta + 1}{2N|\Omega|} \frac{\theta}{j^*} |\Omega| N \frac{1}{p_j} \\
&= (1 - \sigma) \frac{s_j}{p_j} - \frac{\eta\theta}{2 \left(1 - \left(\frac{j}{j^*}\right)^\eta\right)} \frac{s_j}{p_j}.
\end{aligned} \tag{29}$$

Equation (29) implies that product j 's price elasticity of demand ε_j can be written as:

$$\varepsilon_j = 1 - \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j}$$

³⁶In Appendix E we use numerical simulations which do not require any distributional assumption on tastes to verify that the approximation that follows is highly accurate.

$$= \underbrace{\sigma}_{\text{Intensive Margin}} + \underbrace{\left(1 - \left(\frac{j}{j^*}\right)^\eta\right)^{-1} [\theta/2 - (\sigma - 1)]}_{\text{Extensive Margin}} > \sigma. \quad (30)$$

In addition to the standard intensive margin term σ , there is a strictly positive contribution from the extensive margin, since lowering the price of a product can induce new households to start consuming the product. Low j or "mass-market" products are consumed by many households, so the intensive margin is relatively more important for them. High j or "niche" products are consumed by few households, so the extensive margin is relatively more important. As a result, the elasticity of demand increases as market share falls.³⁷ As $j \rightarrow j^*$ and a product approaches the point where it is dropped from the aggregate consumption bundle, the elasticity approaches infinity, i.e. $\varepsilon \rightarrow \infty$. The markup μ_j can then be written (in gross terms) as:

$$\mu_j = \frac{\varepsilon_j}{\varepsilon_j - 1} = \frac{\sigma + \frac{\theta(\eta+1)}{2j^*s_j}}{\sigma + \frac{\theta(\eta+1)}{2j^*s_j} - 1}, \quad (31)$$

and ranges from a high of $(1 + 2/\theta)$ for the largest good $j = 0$ to a low of 1 for $j = j^*$.

The "aggregate markup" is equal to the ratio of aggregate sales to aggregate costs. Using equations (19) and (30), it can be written as:

$$\mu^{Agg} = \frac{\int_0^{j^*} s_j dj}{\int_0^{j^*} s_j \frac{\varepsilon_j - 1}{\varepsilon_j} dj} = \left[\frac{\theta + (\sigma - 1)^2}{\sigma^2} - \frac{1}{2} \frac{\eta\theta^2}{\sigma^2} \left(\frac{\eta + 1}{2 + \theta} \right) \times {}_2F_1 \left(1, \frac{1}{\eta}; 1 + \frac{1}{\eta}; \frac{2\sigma}{2 + \theta} \right) \right]^{-1}, \quad (32)$$

where ${}_2F_1(\cdot)$ is the hypergeometric function.³⁸ Importantly, while this aggregate profit share is a relatively complicated function of σ and θ , it is *not* a function of \tilde{N} , F , or ε .³⁹

4.6 Understanding the Empirical Trends

We now confront our model with concentration measures and other moments from the data to infer which structural forces led to the rise in niche consumption. Collecting previous results, our model

³⁷Interestingly, for good $j = 0$, which has the largest aggregate demand, the positive impact of σ on the elasticity coming through the intensive margin exactly cancels with the negative impact of σ coming from the extensive margin, leaving a total elasticity of $(\theta/2 + 1)$. This result echos a closely related point in [Chaney \(2008\)](#), where the impact of the equivalent parameter for the elasticity of trade flows to trade costs also fully cancels when combining the intensive and extensive margin effects.

³⁸The hypergeometric function is defined as follows: ${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$. $(x)_n$ is the Pochhammer symbol, which equals $\frac{(x+n-1)!}{(x-1)!}$ for all $n > 0$ and equals 1 for $n = 0$.

³⁹We note that if each consumer's taste for each good remains fixed, and markups change for any reason, this would result in a change in the price-adjusted taste distribution and would affect the expressions above that were derived assuming that price-adjusted tastes were distributed a la Pareto. We explore this in more detail in [Appendix E](#), but here note that in order to preserve the Pareto distribution of price-adjusted tastes in the face of increases in N and endogenous markups, the changes in tastes that we additionally require are relatively minor.

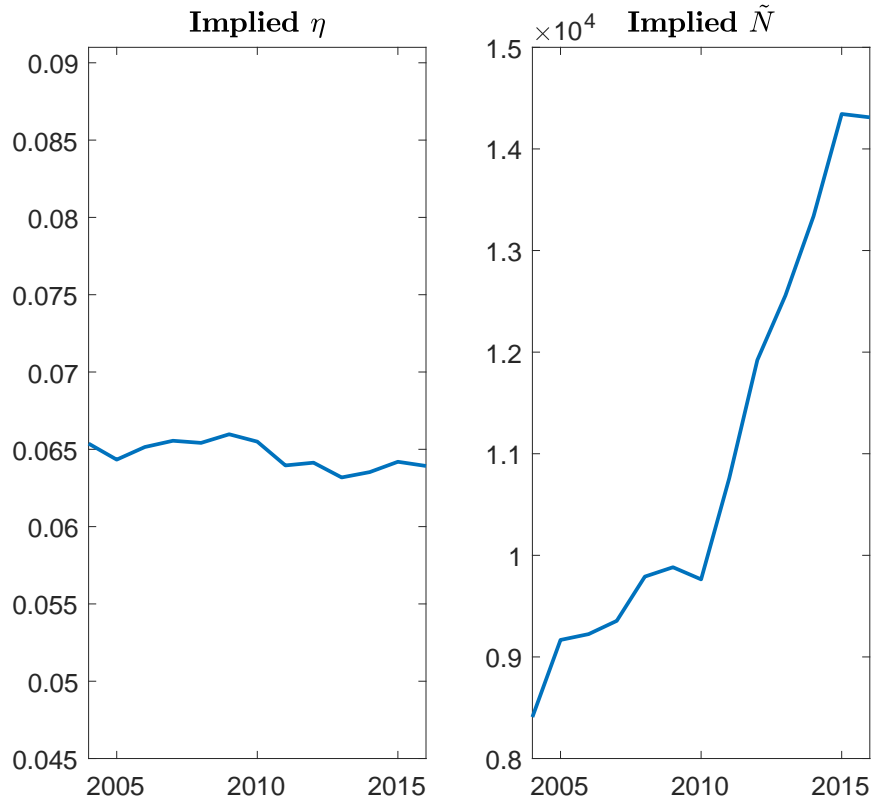
implies that:

$$\mathcal{H}^{\text{HH}} = \frac{(\eta + 1)^2}{4\eta} \frac{1}{|\Omega|} \quad (33)$$

$$\mathcal{H}^{\text{Agg}} = \frac{2(\eta + 1)}{(2\eta + 1)} \left(\frac{1}{2\tilde{N}|\Omega|} \right)^{\frac{1}{2}}. \quad (34)$$

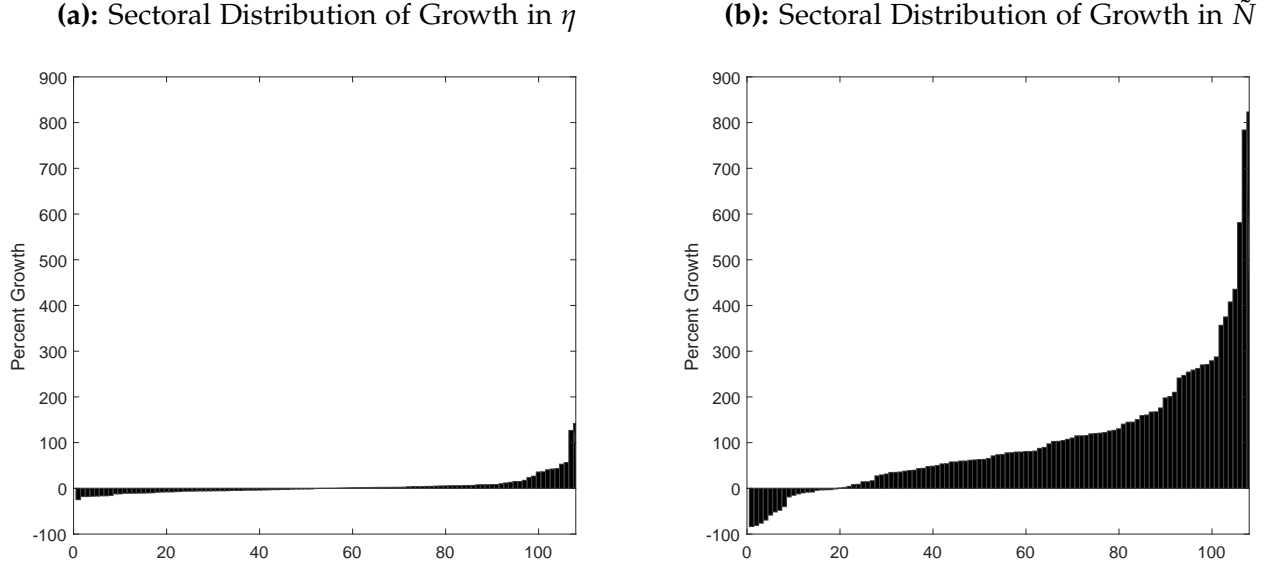
Since \mathcal{H}^{HH} , \mathcal{H}^{Agg} , and $|\Omega|$ are directly observable in the data, this produces a system of two equations – (33) and (34) – that can be solved to determine η and \tilde{N} for each year. Figure 8 shows the time-series for \tilde{N} and η necessary to hit these observables in each year. Through the lens of the model, given the observed path for $|\Omega|$, matching the concentration trends requires nearly constant values for η and a strong upward trend in \tilde{N} . From 2004-2016, η falls by 2 percent while \tilde{N} rises by 70 percent.

Figure 8: Implied Drivers of the Rise of Aggregate Niche Consumption



As above, we can interpret our model as applying at each individual sector and apply equations (33) and (34) to data on concentration trends and varieties consumed sector-by-sector. Rather than generate a single time series for η and a single time series for \tilde{N} for the aggregate economy, as plotted in Figure 8, this allows us to generate time series for these parameters for each sector. Figure 9a plots the distribution of implied growth (in percent terms) from 2004-2016 of η for each sector, and Figure 9b plots the same for the implied growth of \tilde{N} .

Figure 9: Implied Drivers of the Rise of Sector-Level Niche Consumption



It turns out that our aggregate results capture well the sectoral results. Growth in \tilde{N} is essential for explaining the rise in niche consumption even at the sector level, while η typically has not changed. Whereas η declined by 2 percent during 2004-2014 for the aggregate economy, the 25th to 75th percentile of sector-level growth in η over that period ranges from a decline of 7 percent to a rise of 2 percent. Our inferred aggregate \tilde{N} grew by 70 percent, whereas the 25th to 75th percentiles for growth in the sectoral values ranges from 35 percent to 138 percent. How does this cross-sector growth in varieties compare to estimates in the data? As we have emphasized, measuring the total number of varieties purchases in the data can only be done with substantial noise (see Appendix C). However, Figure shows that model implied growth in j^* is broadly in line with the rough measures we can construct in the data, so we think this provides some direct confirmation of the model inference.⁴⁰

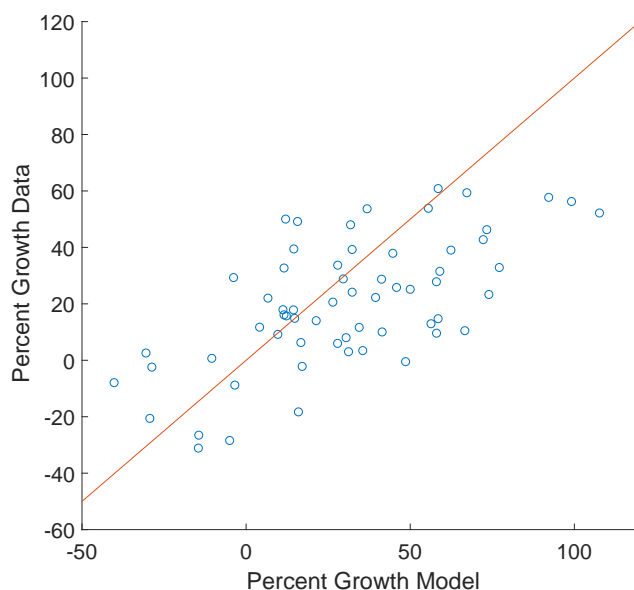
Increases in \tilde{N} can arise from increases in α or N . Changes in α are straightforward to interpret, since α is simply an exogenous parameter governing preference heterogeneity. Our empirical results show that the rise of niche consumption occurs pervasively across all of our narrowly-defined demographic groups. The within-group trends are far more important than across-group trends in generating our aggregate results. While this does not rule out increases in α as a driving force, it seems unlikely that fundamental preferences within narrow groups have become dramatically more heterogeneous over a twelve-year period. Based on this logic, we hold α fixed and interpret increases in \tilde{N} as increases in N in most of our model results.⁴¹

Finally, note that the above exercise takes $|\Omega|$ directly from the data and then solves for the η

⁴⁰This figure constructs statistics excluding products with aggregate spending less than \$100. Using other thresholds changes the relationship some, but there is always a strong positive correlation

⁴¹ N impacts the extent of selection effects while α does not, so if one considers that growth in \tilde{N} is driven in part by growth in α , the welfare gains discussed in the next section will be proportionately smaller.

Figure 10: Consumed Variety (j^*) Growth: Model vs. Data



and N necessary to fit concentration trends. This exercise shows that increases in N are necessary to fit the data, but they are not sufficient. The implied increase in N matches the divergence between household and aggregate concentration given the empirical decline in $|\Omega|$, but as shown in equation (10), increases in N lead to increases in $|\Omega|$. This means that an additional force is required in order to fully fit the empirical trends.

Equation (10) shows that if N increases, declines in $|\Omega|$ must reflect declines in measured real expenditures $\left(E(1 + (1 - \sigma) / \theta)^{\frac{1}{1-\sigma}} b\right)$, increases in effective costs per number of products consumed (F or ϵ), or declines in an "effective curvature" of utility term $\left(\frac{1}{1-\sigma} - \frac{1}{\theta}\right)$.⁴² Expenditures, however, increase in the data, and while changes in either σ or θ could change the curvature term, they would have to change in a very particular way so as to match the decline in $|\Omega|$ without leading to changes in η . We therefore find it most plausible that the decline in $|\Omega|$ reflected an increase in F or ϵ .⁴³ The exact change required to hit the data depends on the particular calibration, but as we show in the next section, this increase in fixed costs has only a modest effect on our welfare conclusions and quantitative conclusions are largely driven by the change in N .

4.7 Implications of Rising Niche Consumption

What are the implications of the rise in niche consumption for welfare and market power? In order to assess this, we set all parameter values of the model to match key empirical moments in 2004, and then

⁴²Real measured expenditures equal E divided by the expression labeled "Average Price" in equation (9).

⁴³While some technological advances such as the rise of the internet or better advertising technology might be expected to lower variety costs, it is also likely that increases in the number of available varieties N make it more costly to sort through and identify the particular products a household wants to purchase. An increase in F or ϵ can be interpreted as a simple proxy for these latter forces when it is accompanied by the increase in N .

look at combinations of N , η and the fixed costs of variety which fit the rise in niche consumption. However, as shown in the ideal price index in equation (9), the welfare implications of this change will depend separately on θ and σ , while our results so far only identify $\eta = 1 - 2(\sigma - 1)/\theta$. Our markup results show that product level elasticities are heterogeneous and are a complicated function of both σ and θ , so that we cannot use existing estimates of the elasticity of demand to identify σ . However, once we identify η , we can use our expression for the aggregate markup in Equation 32 to target specific empirical values and thus separately identify σ and θ . In our baseline results, we target a value for aggregate ratio of revenues to costs of 1.15 in 2004 in order to separately identify σ and θ and implied welfare effects, and then we consider robustness to this target. This implies $\sigma = 4.7$ and $\theta = 7.9$.⁴⁴

We then explore the welfare implications of several different counterfactuals. Given the importance of changes in N identified in the previous section, we begin by increasing N by the amount backed out in Figure 8 while holding all other parameters fixed at their initial 2004 values. We then calculate the resulting change in household welfare, expressed as the percentage change in expenditures on the initial set of goods that would bring the same change in household utility as that delivered by the increase in N . Table 1 shows the resulting change in welfare, expressed as an annual growth rate. We find that a 70 percent increase from N_{2004} to N_{2016} generates total welfare gains of approximately 7 percent, or 0.56 percent per year. That is:

$$U_{2016} = \frac{E}{P_{N_{2016}}} - F \times (|\Omega_{N_{2016}}|)^{\epsilon} = 1.0695 \times \frac{E}{P_{N_{2004}}} - F \times (|\Omega_{N_{2004}}|)^{\epsilon}, \quad (35)$$

where we change N and calculate the endogenous change in P and Ω , but hold fixed all other parameters. We next perform a similar counterfactual but now also including changes in F (or ϵ) to hit also the change in Ω . Implied welfare growth remains large at around 0.46%, so that the effect of increasing fixed costs is quantitatively small relative to the increase in N . Finally, we additionally add changes in σ or θ to hit the small decline in η in Figure 8, and this has almost no effect on the results. These welfare gains come from a variety of sources, which we also decompose in Table 1. The most quantitatively important effect comes from the third term in the ideal price index in equation (9): an increase in N for a given $|\Omega|$ leads to gains from selection. With more choices, households consume those products better suited to their particular tastes. Second, changes in $|\Omega|$ show up as standard love-for-variety effects on welfare, even if selection effects $|\Omega|/N$ are held constant. Third, changes in $|\Omega|$, F or ϵ lead to changes in the fixed costs paid by households. Finally, in specifications where we change σ or θ , there is a mild change in the “average price” effect which just captures the nominal price of one unit of the consumption bundle.⁴⁵ Overall, the conclusions from Table 1 are simple: the rise of niche consumption is associated with substantial welfare gains, and these arise almost entirely

⁴⁴Other parameters are not important for our quantitative conclusions, but we set $\alpha = 0.36$, $\epsilon = 2$, $E = 35$, $b = 1$, $F = 0.055$. E is set to match average household category expenditures. Given b and ϵ , we choose F to match Ω .

⁴⁵Note that because of non-linearities, this decomposition need not add exactly to the overall effect.

from greater selection as N increases.

It is useful to note that if one simply viewed our data through the lens of a representative household model with CES preferences, it might be natural to only consider the love-of-variety loss in column 3 and misleadingly conclude that welfare *declined* from 2004-2016, since the typical household consumes fewer varieties in 2016 than in 2004. Heterogeneity in product consumption across households is crucial for capturing the divergent concentration trends in our data. Representative agent models abstract from this heterogeneity, and our results show that this can potentially lead to misleading conclusions about the welfare effects arising from changes in the number of products households consume. Alternatively, one could calibrate a representative agent model purely to aggregate variety counts. However, as we emphasize in Appendix C, aggregate variety growth cannot be measured with much precision so this exercise produces a huge range of potential conclusions.⁴⁶

Table 1: Annualized Welfare Growth (Compensating Expenditures)

	Total $d \ln E$	Ave Price $\left(\frac{1+\theta-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}}$	Love-of-Variety $(\Omega)^{\frac{1}{1-\sigma}}$	Selection $(\Omega /N)^{\frac{1}{\theta}}$	Fixed Costs $-F \times (\Omega)^{\epsilon}$
$\uparrow N$	0.56%	0%	0.08%	0.52%	-0.05%
$\uparrow N, \uparrow \epsilon$	0.47%	0%	-0.13%	0.62%	-0.02%
$\uparrow N, \uparrow F$	0.46%	0%	-0.13%	0.62%	-0.04%
$\uparrow N, \uparrow F, \uparrow \sigma$	0.45%	-0.001%	-0.14%	0.62%	-0.03%
$\uparrow N, \uparrow F, \downarrow \theta$	0.47%	-0.003%	-0.13%	0.63%	-0.04%

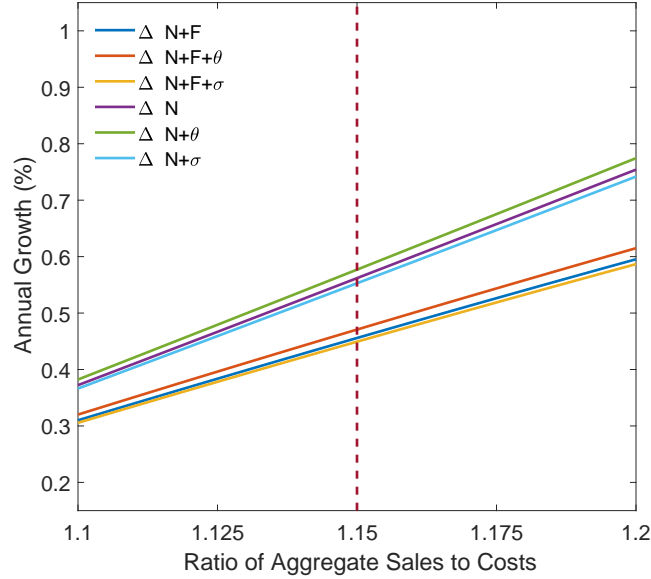
The welfare effects of increased product selection do depend importantly on θ , which as noted above, we pin down by targeting an aggregate markup of 1.15 together with the initial η value in Figure 8. Figure 11 shows how implied welfare growth changes under alternative calibrations for the initial aggregate markup. The dashed red line corresponds to our baseline calibration and the overall numbers in the first column of Table 11. This figure shows that welfare effects remain large for a wide range of markup calibration choices, which pin down θ and σ given the value of η .

What are the sectoral implications of the rise in niche consumption? As above, we do this identical calibration and counterfactual exercise at the sector level and report these results in Table 3.⁴⁷ Cate-

⁴⁶Substituting aggregate variety counts under the different measurement assumptions in Table A2 into a representative agent model with $F = 0$ (so $|\Omega| = N$), and using the same baseline $\sigma = 4.7$ implies a huge range in annual welfare growth from 0.1% to 0.57%. In contrast Table A3 shows that our micro data based inference is almost totally insensitive to these measurement issues.

⁴⁷We set the initial $|\Omega|$ to fit the empirical value for each value and given this $|\Omega|$, we set η to match each sector's \mathcal{H}^{HH} . Given the sector's η , we pick the separate values of θ and σ to generate a ratio of total revenues to total costs of 1.15 (see equation 32 below). If we instead impose a common σ for each sector to pin down σ and θ rather than targeting a common

Figure 11: Robustness of Welfare Calculations to Calibration of θ and σ



gories such as coffee, snacks, and soup all rank in the top-10 for gains from the increase in the number of products. Not all sectors exhibit such gains, though. For example, the number of imputed product varieties declines, and brings welfare losses for eggs, cottage cheese and frozen vegetables.

4.8 Implications for Markups and Aggregate Profits

How large are changes in aggregate market power arising from the rise of niche consumption? It turns out they are extremely small. As can be seen in expression 32, aggregate market power does not vary at all with changes in N , F , or ϵ . These parameters have implications for the distribution of markups across products in the economy, since they impact j^* , but changing them leaves the aggregate markup exactly constant, since j^* drops out of equation (32). More intuitively, this result arises from two opposing forces which exactly cancel when j^* changes. On the one hand, the j th good in an economy with a low j^* is closer to being the marginal consumed good and will therefore have a lower markup than the j th good in an economy with a high j^* . This can be seen in equation (30), which shows that the elasticity of demand is strictly increasing in the ratio of j to j^* . All else equal, this selection force raises aggregate markups. On the other hand, in an economy with greater product choice, the high-markup products account for a smaller share of aggregate spending. This competition force from increasing N reduces aggregate markups. Equation 32 shows that in the aggregate, these opposing forces exactly cancel and the ratio of total revenues to total costs remains unchanged. In specifications where we also change σ or θ , aggregate markups are no longer exactly fixed but resulting changes are extremely small (rising by 0.02 percentage points if we vary θ to hit the change in η and falling by

ratio of revenues to costs, we find similar welfare conclusions. To save space, and to minimize the influence of measurement error, we exclude small categories which account for less than 0.5 percent of all spending.

Table 2: Sectoral Welfare Growth Associated with Rise in Niche Consumption, 2004-2016

		Annual % ΔU with:			Annual % ΔU with:			
		ΔN	$\Delta N, \Delta\theta, \Delta F$	$\Delta N, \Delta\sigma, \Delta F$	ΔN	$\Delta N, \Delta\theta, \Delta F$	$\Delta N, \Delta\sigma, \Delta F$	
(1) Coffee		2.37	3.31	2.65	(33) Frozen Meats and Seafood	0.62	0.56	0.56
(2) Disposable Diapers		1.66	1.13	1.28	(34) Tea	0.61	1.11	0.76
(3) Snacks		1.4	1.47	1.37	(35) Canned Vegetables	0.59	0.44	0.34
(4) Prepared Deli Foods		1.39	1.44	1.33	(36) Vitamins	0.58	0.68	0.59
(5) Pet Food		1.37	0.99	1.02	(37) Beer	0.57	0.97	0.72
(6) Skin Care Preparations		1.36	1.26	1.21	(38) Pet Care	0.52	0.68	0.55
(7) Soup		1.33	1.07	1.04	(39) Hair Care	0.51	0.32	0.32
(8) Detergents		1.3	0.89	1.09	(40) Nuts	0.49	0.42	0.36
(9) Breakfast Food		1.21	0.96	1.01	(41) Crackers	0.42	0.42	0.3
(10) Pizza		1.15	0.78	0.94	(42) Cosmetics	0.41	0.92	0.5
(11) Carbonated Beverages		1.08	0.9	0.88	(43) Ice Cream	0.38	0.25	0.18
(12) Oral Hygiene		1.03	1.01	0.8	(44) Ready-to-Serve Foods	0.38	0.07	0.12
(13) Canned and Bottled Juice Drinks		1.01	0.71	0.74	(45) Household Cleaners	0.35	0.37	0.08
(14) Light Bulbs and Electric Goods		1	-0.62	0.5	(46) Wine	0.35	0.43	0.35
(15) Household Supplies		0.94	0.39	0.56	(47) Packaged Deli Meats	0.33	0.21	0.22
(16) Housewares and Appliances		0.94	0.01	0.57	(48) Desserts, Gelatins, and Syrup	0.32	0.15	0.09
(17) Personal Soap And Bath Additives		0.88	1.4	0.86	(49) Prepared Foods (dry mixes)	0.31	0.25	0.15
(18) Cookies		0.85	0.82	0.73	(50) Baking Supplies	0.31	0.31	0.15
(19) Condiments and Gravies		0.84	0.84	0.74	(51) Stationary and School Supplies	0.29	-0.32	-0.19
(20) Butter and Margarine		0.82	0.35	0.6	(52) Cough and Cold Remedies	0.14	0.21	0.11
(21) Cereal		0.81	0.54	0.57	(53) Spices, Seasonings, and Extracts	0.09	0.35	0.16
(22) Liquor		0.81	0.73	0.75	(54) Tobacco	0.01	-0.17	-0.11
(23) Jams and Jellies		0.78	0.82	0.65	(55) Packaged Milk	-0.01	0.27	0.08
(24) Medications		0.77	0.85	0.64	(56) Records and Tapes	-0.03	-1.43	-0.49
(25) Yogurt		0.76	0.97	0.72	(57) Salad Dressings and Mayonnaise	-0.09	-0.22	-0.3
(26) Laundry Supplies		0.74	0.58	0.42	(58) Frozen Vegetables	-0.14	-0.02	-0.19
(27) Cheese		0.74	0.95	0.82	(59) Non-Carbonated Soft Drinks	-0.17	-0.05	-0.15
(28) Batteries and Flashlights		0.71	0.08	0.22	(60) Paper Products	-0.68	-0.82	-0.71
(29) Candy		0.66	0.63	0.44	(61) Bread and Baked Goods	-0.75	-0.83	-0.79
(30) Prepared Frozen Foods		0.63	0.44	0.46	(62) Eggs	-0.91	-0.73	-0.79
(31) Milk (non-packaged)		0.63	0.69	0.59	(63) Cottage Cheese and Sour Cream	-1.22	-1.23	-1.14
(32) Wrapping Materials And Bags		0.62	0.3	0.29	(64) Photographic Supplies	-1.83	-6.48	-3.07

Note: Table shows welfare changes (in percent), as defined in equation (35), when equations (33) and (34) are used to back out key parameters from the data at the sector level. For each sector we report three values. The column labeled with “ ΔN ” corresponds to the welfare change when vary only N , to match the divergence between \mathcal{H}^{HH} and \mathcal{H}^{Agg} for each sector, but not the individual changes in \mathcal{H}^{HH} , \mathcal{H}^{Agg} or $|\Omega|$. The column labeled “ $\Delta N, \Delta F, \Delta\theta$ ” additionally varies the per-variety fixed cost parameter and varies η by changing θ to exactly hit the change in \mathcal{H}^{HH} , \mathcal{H}^{Agg} and $|\Omega|$. The column labeled “ $\Delta N, \Delta F, \Delta\sigma$ ” similarly targets the changes in all three moments, but varies η by changing σ instead of θ .

0.003 percentage points if we instead vary σ to hit the change in η).

Thus, even though markups are endogenous in our model and there are large diverging concentration trends, the rise of niche consumption in our model is associated with essentially no change in aggregate market power. Our model therefore shows how the economy can exhibit large changes in aggregate and household concentration without any change in aggregate market power.

Finally, Table 3 shows the corresponding changes in market power at the individual sector level when we again re-estimate our model sector by sector. We compute this by assuming that sector level changes in η are driven either entirely by θ (Column 1) or σ (Column 2). Overall, the conclusions mirror that from the aggregate analysis, that the typical sector has essentially no change in aggregate markups, however there are a few sectors with non-trivial changes. Concentrating on the larger changes induced by θ variation in Column 1, there are non-trivial declines in markups in photographic supplies, records and tapes and lightbulbs and non-trivial increases in coffee, soaps and bath additives,

tea, and cosmetics.⁴⁸

Table 3: Markup Changes Associated with Rise in Niche Consumption, 2004-2016

		Percentage Point $\Delta\mu^{Agg}$ with:		Percentage Point $\Delta\mu^{Agg}$ with:			
		$\Delta N, \Delta\theta, \Delta F$	$\Delta N, \Delta\sigma, \Delta F$	$\Delta N, \Delta\theta, \Delta F$	$\Delta N, \Delta\sigma, \Delta F$		
(1)	Photographic Supplies	-4.68	0.47	(33)	Eggs	0.05	-0.01
(2)	Records and Tapes	-0.89	0.09	(34)	Wine	0.05	-0.01
(3)	Light Bulbs and Electric Goods	-0.79	0.1	(35)	Vitamins	0.06	-0.01
(4)	Housewares and Appliances	-0.38	0.05	(36)	Snacks	0.06	-0.01
(5)	Butter and Margarine	-0.22	0.03	(37)	Milk (non-packaged)	0.06	-0.01
(6)	Detergents	-0.14	0.02	(38)	Cookies	0.06	-0.01
(7)	Household Supplies	-0.13	0.02	(39)	Canned Vegetables	0.07	-0.01
(8)	Batteries and Flashlights	-0.12	0.01	(40)	Condiments and Gravies	0.07	-0.01
(9)	Pizza	-0.12	0.01	(41)	Cough and Cold Remedies	0.07	-0.01
(10)	Disposable Diapers	-0.11	0.01	(42)	Prepared Deli Foods	0.07	-0.01
(11)	Paper Products	-0.09	0.01	(43)	Salad Dressings and Mayonnaise	0.07	-0.01
(12)	Stationary and School Supplies	-0.08	0.01	(44)	Prepared Foods (dry mixes)	0.08	-0.01
(13)	Cottage Cheese and Sour Cream	-0.08	0.01	(45)	Non-Carbonated Soft Drinks	0.08	-0.01
(14)	Tobacco	-0.05	0.01	(46)	Cheese	0.09	-0.01
(15)	Ready-to-Serve Foods	-0.03	0	(47)	Pet Care	0.09	-0.01
(16)	Breakfast Food	-0.03	0	(48)	Crackers	0.1	-0.01
(17)	Pet Food	-0.03	0	(49)	Candy	0.12	-0.02
(18)	Bread and Baked Goods	-0.03	0	(50)	Laundry Supplies	0.12	-0.01
(19)	Canned and Bottled Juice Drinks	-0.02	0	(51)	Frozen Vegetables	0.12	-0.02
(20)	Cereal	-0.02	0	(52)	Baking Supplies	0.12	-0.02
(21)	Liquor	-0.02	0	(53)	Spices, Seasonings, and Extracts	0.13	-0.02
(22)	Prepared Frozen Foods	-0.01	0	(54)	Jams and Jellies	0.13	-0.02
(23)	Packaged Deli Meats	-0.01	0	(55)	Medications	0.14	-0.02
(24)	Hair Care	0	0	(56)	Oral Hygiene	0.15	-0.02
(25)	Frozen Meats and Seafood	0	0	(57)	Packaged Milk	0.16	-0.02
(26)	Wrapping Materials And Bags	0.01	0	(58)	Yogurt	0.18	-0.02
(27)	Carbonated Beverages	0.02	0	(59)	Beer	0.2	-0.03
(28)	Soup	0.02	0	(60)	Household Cleaners	0.22	-0.03
(29)	Skin Care Preparations	0.03	0	(61)	Cosmetics	0.25	-0.03
(30)	Desserts, Gelatins, and Syrup	0.05	-0.01	(62)	Tea	0.25	-0.03
(31)	Ice Cream	0.05	-0.01	(63)	Personal Soap And Bath Additives	0.37	-0.05
(32)	Nuts	0.05	-0.01	(64)	Coffee	0.46	-0.06

5 Conclusions

This paper empirically documents a rise in what we call "niche" consumption. Households are increasingly concentrating their spending. This pattern, however, does not appear to be driven by the emergence of superstar products. Rather, households are increasingly buying different goods from one another. The increase in segmentation seen in many other walks of modern life also applies to consumption: our grocery baskets look less and less similar. As a result, aggregate spending has become less concentrated.

We develop a new model of product demand in order to explore the drivers and implications of the rise in niche consumption. In our model, households choose how many products to consume, spend different amounts on each good, and differ from other households in their choice of which products to buy. The model delivers simple analytical expressions for household and aggregate concentration

⁴⁸The aggregate markup is more sensitive to θ than to σ , so this has a more significant effect. Holding η fixed, θ and σ move markups in opposite directions, so it is not surprising that the results in the two columns are perfectly correlated.

indices, and these closed form solutions allow us to match the model to data and infer the drivers of our empirical findings. Increases in product availability played a critical role in the divergent concentration trends, and led to welfare gains from households being able to consume a subset of products that better satisfied their tastes. This welfare effect would be difficult to find using standard statistics such as measured price indices. Finally, our model delivers endogenous and heterogeneous markups. Matching the trends in household and aggregate concentration carries implications for the distribution of markups, but does not imply changes in aggregate market power.

Our model highlights the importance of greater product choice but treats the set of available products as an exogenous parameter. We suspect the nature of product introduction and development, however, reflects recent progress in supply chain integration, big-data marketing research, targeted advertising, and the growing importance of online sales. Unpacking the product innovation process and relating it to these important trends is a fruitful avenue for future research on consumption behavior.

References

- (2019): "Bottom-up Markup Fluctuations," *mimeo*.
- AGUADO, J. M., C. FEIJOO, AND I. MARTINEZ (2015): *Emerging Perspectives on the Mobile Content Evolution*. IGI Global.
- ALLCOTT, H., R. DIAMOND, J.-P. DUBE, J. HANDBURY, I. RAHKOVSKY, AND M. SCHNELL (2017): "Food Deserts and the Causes of Nutritional Inequality," Working Paper 24094, National Bureau of Economic Research.
- 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.
- ANDERSON, E., S. REBELO, AND A. WONG (2018): "Markups Across Space and Time," Working Paper 24434, National Bureau of Economic Research.
- ANDERSON, S., A. DE PALMA, AND J. THISSE (1987): "The CES is a discrete choice model?," *Economics Letters*, 24(2), 139–140.
- ARGENTE, D., M. LEE, AND S. MOREIRA (2018a): "How do firms grow? The life cycle of products matters," .
- (2018b): "Innovation and product reallocation in the great recession," *Journal of Monetary Economics*, 93, 1–20.
- ARKOLAKIS, C., S. DEMIDOVA, P. J. KLENOW, AND A. RODRIGUEZ-CLARE (2008): "Endogenous variety and the gains from trade," *American Economic Review*, 98(2), 444–50.
- ATKESON, A., AND A. BURSTEIN (2008): "Pricing-to-market, trade costs, and international relative prices," *American Economic Review*, 98(5), 1998–2031.
- ATKIN, D., B. FABER, AND M. GONZALEZ-NAVARRO (2018): "Retail globalization and household welfare: Evidence from Mexico," *Journal of Political Economy*, 126(1), 1–73.
- AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. 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).
- BERGER, D., AND J. VAVRA (2015): "Consumption Dynamics During Recessions," *Econometrica*, 83(1).
- BERGER, D., AND J. VAVRA (2019): "Shocks vs. Responsiveness: What Drives Time-Varying Dispersion?," *Journal of Political Economy*.
- BERRY, S. T., M. GAYNOR, AND F. S. MORTON (2019): "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," Working Paper 26007, National Bureau of Economic Research.
- BERTRAND, M., AND E. KAMENICA (2018): "Coming apart? Lives of the Rich and the Poor Over Time," Discussion paper.
- BORNSTEIN, G. (2018): "Entry and Profits in an Aging Economy: The Role of Consumer Inertia," .

- BRODA, C., AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121(2), 541–585.
- BRODA, C., AND D. W. WEINSTEIN (2004): "Variety growth and world welfare," *American Economic Review*, 94(2), 139–144.
- BRYNJOLFSSON, E., Y. HU, AND M. D. SMITH (2003): "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, 49(11), 1580–1596.
- CHANEY, T. (2008): "Distorted gravity: the intensive and extensive margins of international trade," *American Economic Review*, 98(4), 1707–21.
- COIBION, O., Y. GORODNICHENKO, AND D. KOUSTAS (2017): "Consumption Inequality and the Frequency of Purchases," Discussion paper.
- DAVIS, D. R., J. I. DINGEL, J. MONRAS, AND E. MORALES (2017): "How Segregated is Urban Consumption?," .
- DE LOECKER, J., AND J. EECKHOUT (2017): "The Rise of Market Power and the Macroeconomic Implications," Working Paper 23687, National Bureau of Economic Research.
- DUBE, J.-P., G. HITSCH, AND P. ROSSI (2018): "Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession," *Marketing Science*.
- EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2018): "How costly are markups?," Discussion paper, National Bureau of Economic Research.
- FEENSTRA, R. C. (1994): "New product varieties and the measurement of international prices," *The American Economic Review*, pp. 157–177.
- FEENSTRA, R. C., AND D. E. WEINSTEIN (2017): "Globalization, markups, and US welfare," *Journal of Political Economy*, 125(4), 1040–1074.
- FURMAN, J., AND P. ORSZAG (2015): "A firm-level perspective on the role of rents in the rise in inequality," *Presentation at AIAA Just Society Centennial Event in Honor of Joseph Stiglitz Columbia University*, 16.
- 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.
- GOPINATH, G., AND O. ITSKHOKI (2010): "Frequency of price adjustment and pass-through," *The Quarterly Journal of Economics*, 125(2), 675–727.
- HANDBURY, J., AND D. E. WEINSTEIN (2014): "Goods prices and availability in cities," *The Review of Economic Studies*, 82(1), 258–296.
- HANSEN, K., AND V. SINGH (2015): "Choice Concentration," Discussion paper.
- HAUSMAN, J. A. (1996): *Valuation of New Goods under Perfect and Imperfect Competition* pp. 207–248. University of Chicago Press.
- HOTTMAN, C. J., S. J. REDDING, AND D. E. WEINSTEIN (2016): "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 131(3), 1291–1364.
- JARAVEL, X. (2019): "The Unequal Gains from Product Innovations: Evidence from the US Retail Sector," *Quarterly Journal of Economics*, 134(2).

- KIMBALL, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27(4), 1241–1277.
- KROFT, K., J.-W. P. LALIBERTÉ, R. LEAL-VIZCAÍNO, AND M. J. NOTOWIDIGDO (2017): "Quantifying the Welfare Gains of Variety: A Sufficient Statistics Approach," Discussion paper, Mimeo, available at [http://korykroft.com/wordpress/KKLN variety july2017. pdf](http://korykroft.com/wordpress/KKLN%20variety%20july2017.pdf).
- LEVIN, A., AND T. YUN (2008): "Reconsidering the microeconomic foundations of price-setting behavior," *Unpublished Manuscript, Board of Governors of the Federal Reserve System*.
- LI, N. (2018): "An Engel curve for variety," *Working Paper*.
- MELITZ, M. J., AND G. I. OTTAVIANO (2008): "Market size, trade, and productivity," *The review of economic studies*, 75(1), 295–316.
- MICHELACCI, C., L. PACIELLO, AND A. POZZI (2019): "The Extensive Margin of Aggregate Consumption Demand," *Mimeo*.
- NEIMAN, B., AND J. VAVRA (2018): "The Rise of Household Spending Concentration," *Mimeo*.
- PETRIN, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of political Economy*, 110(4), 705–729.
- 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.
- QUAN, T. W., AND K. R. WILLIAMS (2018): "Product variety, across-market demand heterogeneity, and the value of online retail," *The RAND Journal of Economics*, 49(4), 877–913.
- REDDING, S. J., AND D. E. WEINSTEIN (2016): "Measuring aggregate price indexes with demand shocks: Theory and evidence for CES Preferences," Discussion paper, National Bureau of Economic Research.
- ROSSI-HANSBERG, E., P.-D. SARTE, AND N. TRACHTER (2018): "Diverging trends in national and local concentration," Discussion paper, National Bureau of Economic Research.
- STROEBEL, J., AND J. VAVRA (2019): "House Prices, Local Demand and Retail Prices," *Journal of Political Economy*, 127(3).
- SYVERSON, C. (2018): "Remarks at 'Changing Market Structure and Implications for Monetary Policy'," *Remarks, 2018 Jackson Hole Symposium*.

Appendix A. Detailed Data Description

Our primary data set is the AC Nielsen Homescan data, which we use to measure household-level shopping behavior.¹ As discussed in the text, our panel contains weekly household-level product spending for the period 2004-2016. The panel has large coverage, with roughly 170,000 households in over 22,000 zip codes recording prices for almost 700 million unique transactions covering 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.² 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 and scan the barcodes of all purchased items. 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 our baseline analysis, we define a product as a UPC. UPCs are directly assigned by the manufacturer and will typically change any time there is any change in product characteristics. However, we also compute results instead defining a product as a "brand". Information on brands is constructed by Kilts/Nielsen and is 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. We focus on UPCs as our baseline product definition for several reasons: 1) Most importantly, UPCs are directly assigned by the manufacturer, while the brand variable is constructed by Kilts/Nielsen. Which UPCs are grouped into more aggregate brands involves some subjective judgment, and this aggregation is not necessarily consistent across categories or time.

¹These 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.

²We use these projection weights in all reported results, but our results are similar when weighting households equally.

2) UPCs are the most fine-grained definition available and will capture relevant product changes like the introduction of new flavors which will typically not be captured with the brand-definition. 3) In order to preserve anonymity of the stores in the Nielsen sample, all generic UPCs are assigned the same brand code. This means that analysis of brand-level spending can only be done on the subset of name-brand products and must exclude the large and growing share of generic products from the sample. (see e.g. [Dube, Hitsch, and Rossi \(2018\)](#)).

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 or packaging for otherwise equivalent UPCs) as identical products.³ For this reason, we show robustness 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. There is substantial heterogeneity across product modules in the degree of household concentration, so 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 [A2](#) shows that aggregate spending growth in our sample tracks government data on aggregate spending growth in comparable categories. Our conclusions are even stronger when instead using a balanced panel of households to eliminate household composition changes.

While our baseline sample includes all UPCs, we also show that our results hold when excluding generic/private-label products. In order to preserve anonymity of the stores in the Nielsen sample, the exact identity of generic brands in the Nielsen data is masked. 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 show that excluding generics and calculating concentration trends for branded products produces nearly identical results.

Finally, it is also useful to discuss the potential role of online shopping for our measurement.

³It 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 homogeneous across time, using UPCs as our notion of product may lead to spurious changes in concentration with no substantive change in household behavior.

Households in the Nielsen Homescan sample are supposed to scan barcoded purchases of purchases from online retailers in addition to the items they scan from brick-and-mortar retailers. Indeed the Nielsen panel shows a growing share of online spending across time (Figure A1). However, for the categories covered in Nielsen data, online spending is relatively unimportant, so even by the end of the sample these spending shares remain low.⁴ Breaking results out further for particular categories where online spending is likely to be more and less relevant delivers no obvious interaction with concentration trends. For these reasons, we conclude that online shopping is unlikely to be of direct importance for understanding the diverging trends that we document.

Figure A1: Online Spending Shares



⁴Online vs. brick-and-mortar spending is classified at the level of the retail chain. This means that our measure captures spending at online only retailers such as Amazon but misses online spending associated with traditional retailers such as spending at Walmart.com.

Appendix B. Additional Empirical Results

Figure A2: Household Spending in Nielsen vs. Consumer Expenditure Survey

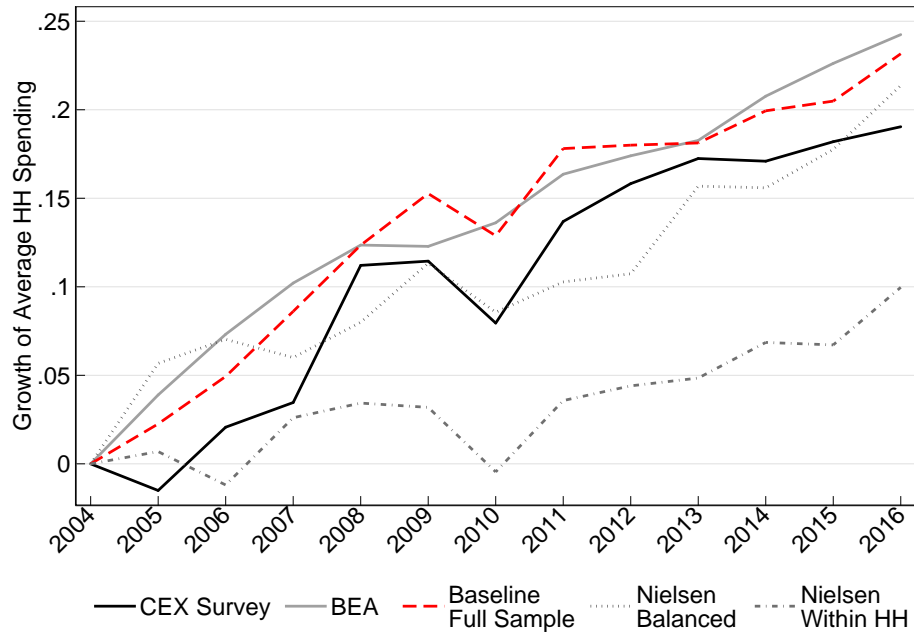


Figure A3: Concentration Trends: Excluding Generics

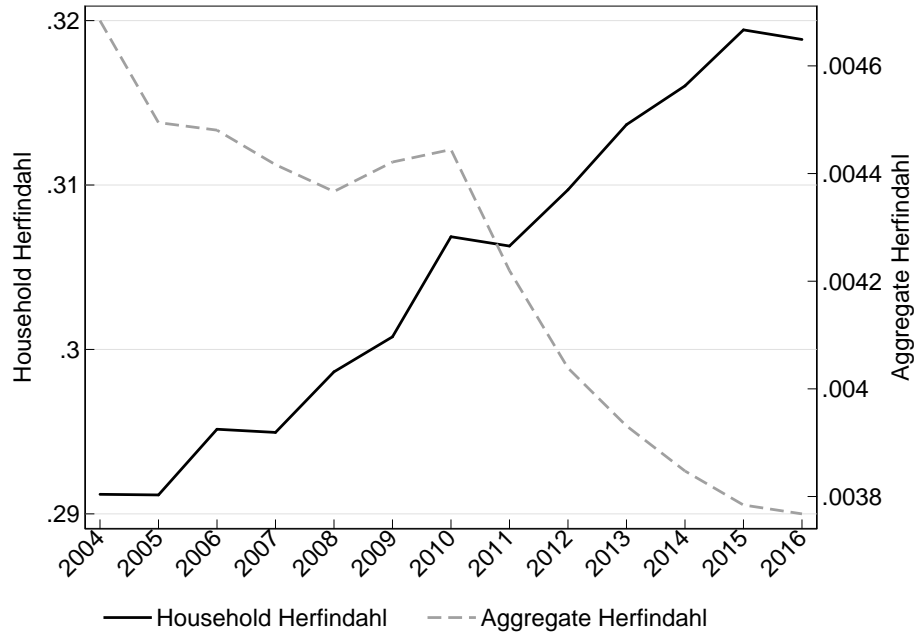


Figure A4: Concentration Trends: Including Category Composition Changes

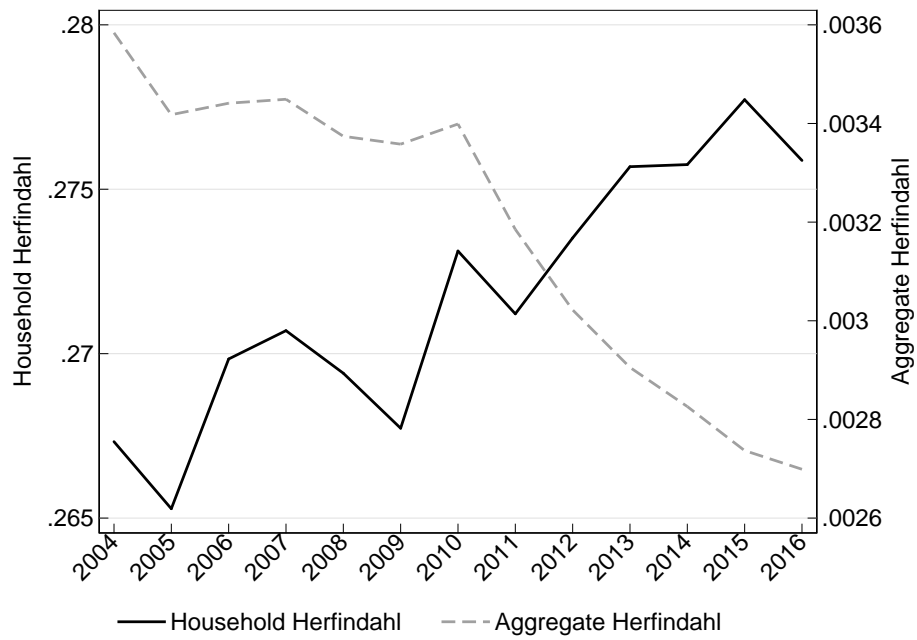


Figure A5: Concentration Trends: Brand Instead of UPC

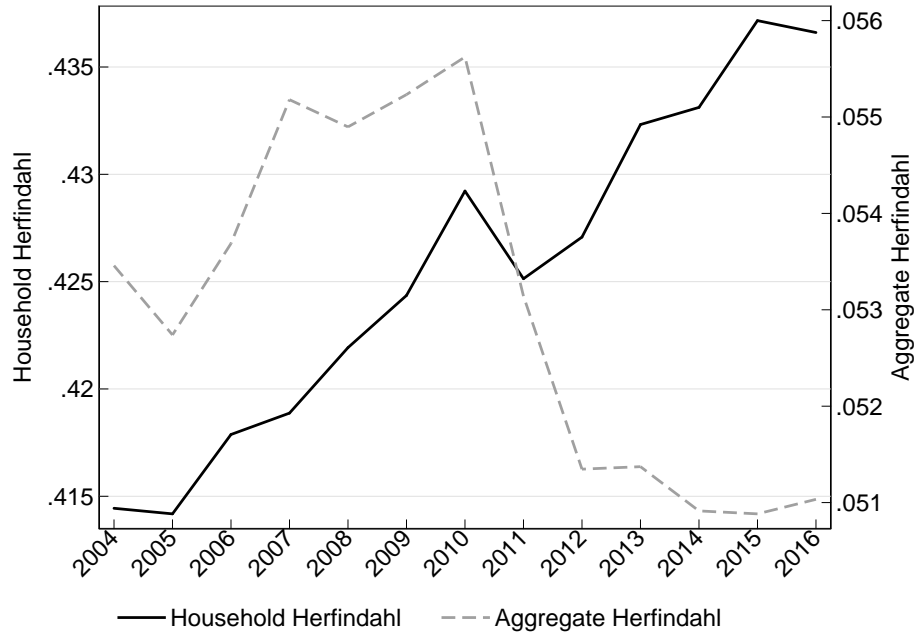


Figure A6: Concentration Trends: Product Module instead of Group

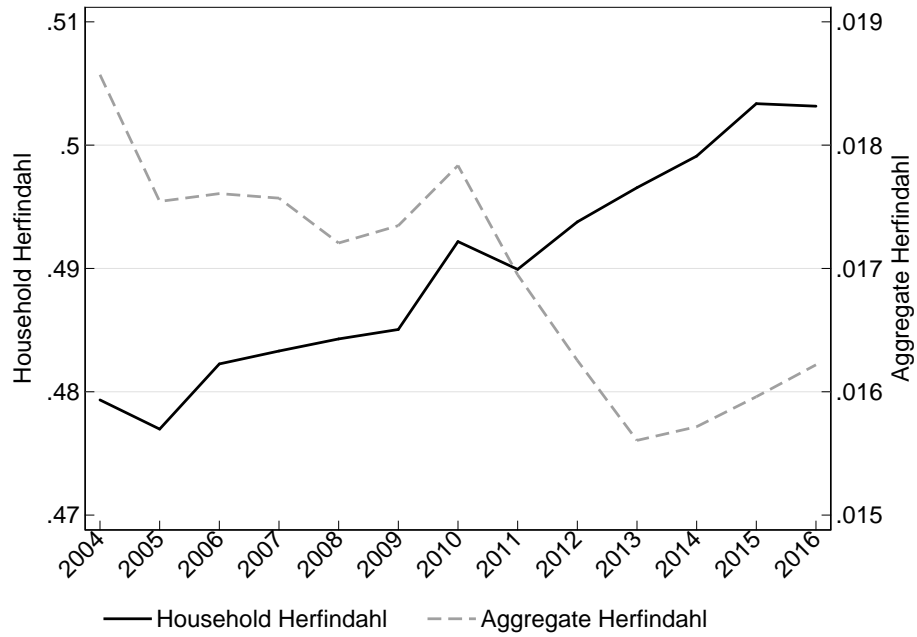


Figure A7: Alternative Concentration Measures

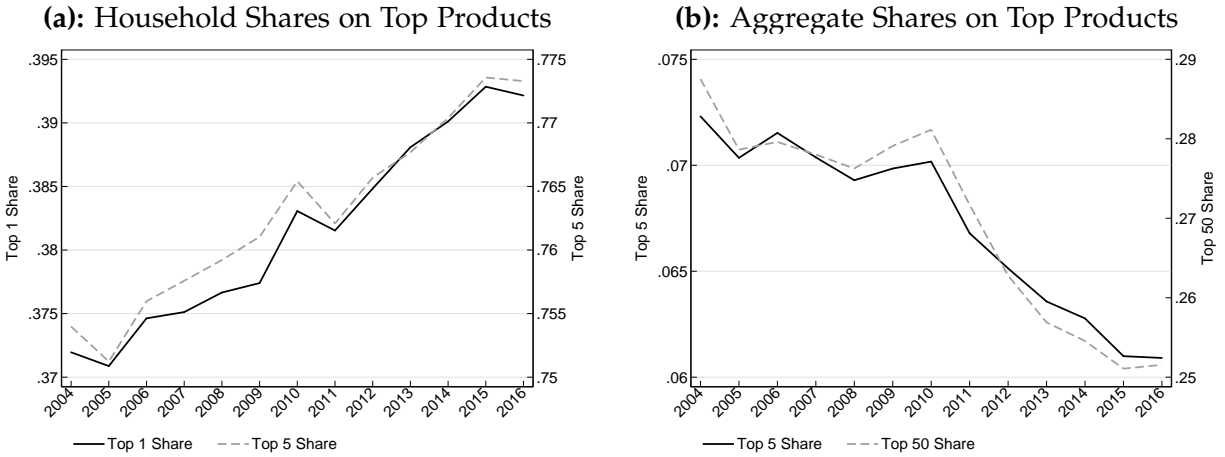


Figure A8: Concentration Trends for Different Samples

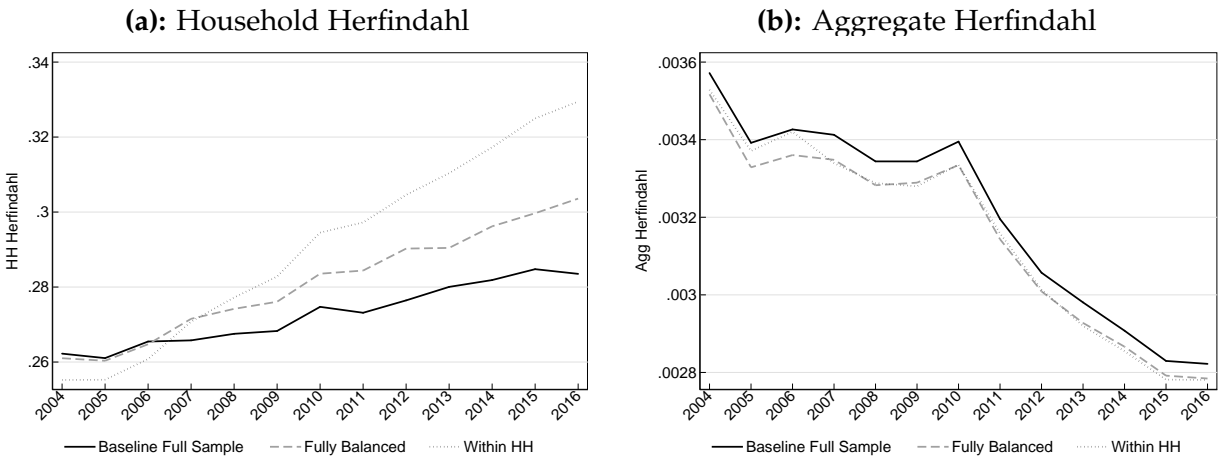


Figure A9: 2004-2016 Concentration Growth by Household Size

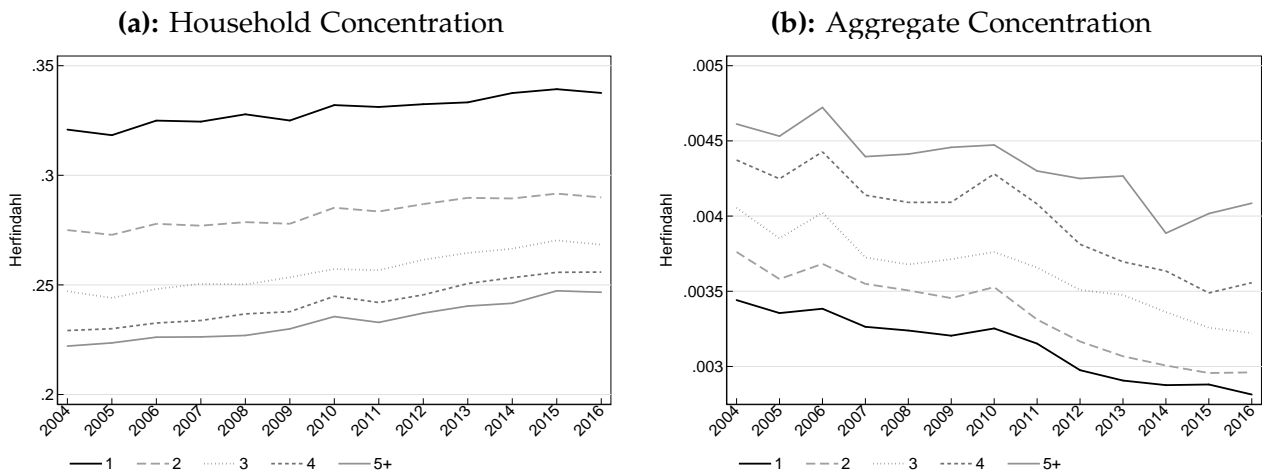
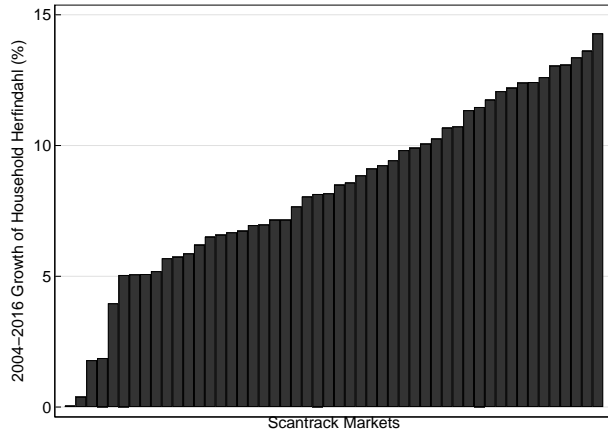


Figure A10: 2004-2016 Concentration Growth Within Location

(a): Household Concentration



(b): Aggregate Concentration

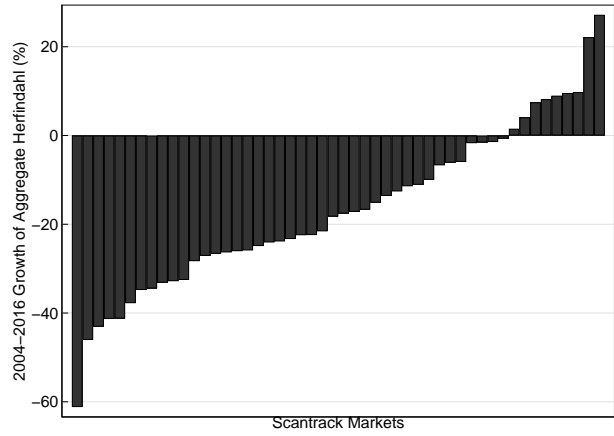
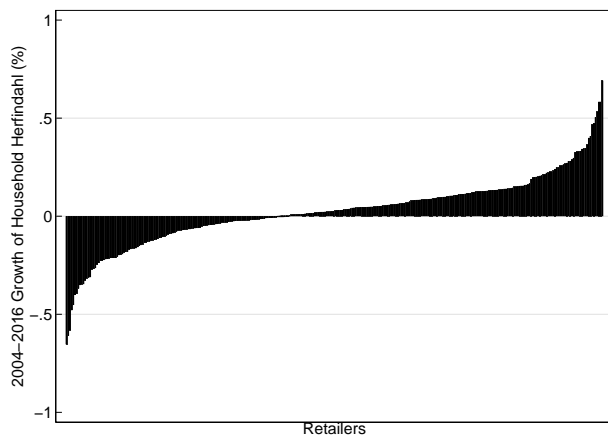


Figure A11: 2004-2016 Concentration Growth Within Retailer

(a): Household Concentration



(b): Aggregate Concentration

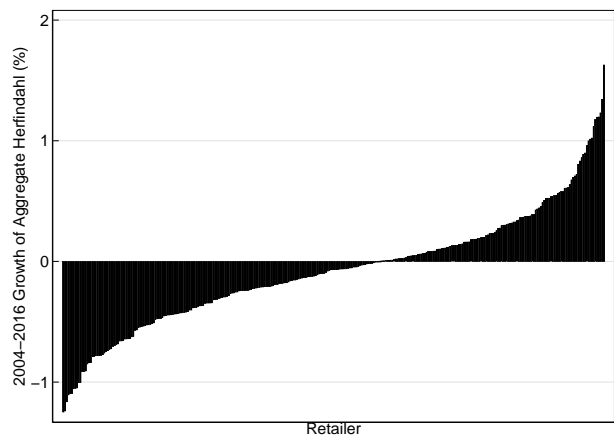


Figure A12: Intensive Margin P v. Q effects for UPCs

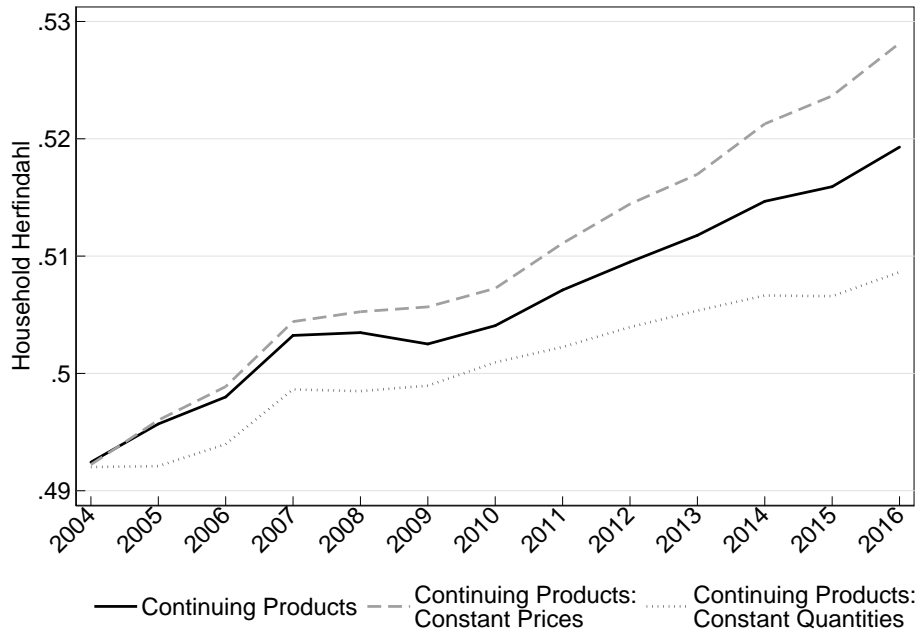


Figure A13: Effects of $\tilde{N} \uparrow$ and $F \uparrow$ on Profit Distribution

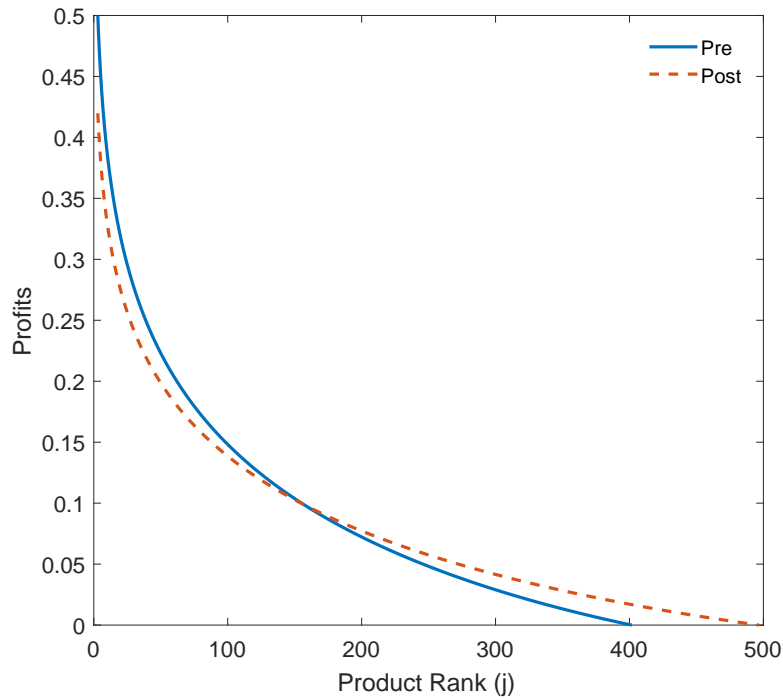


Figure A14: 2004-2016 Concentration growth for continuing vs. all products (brands)

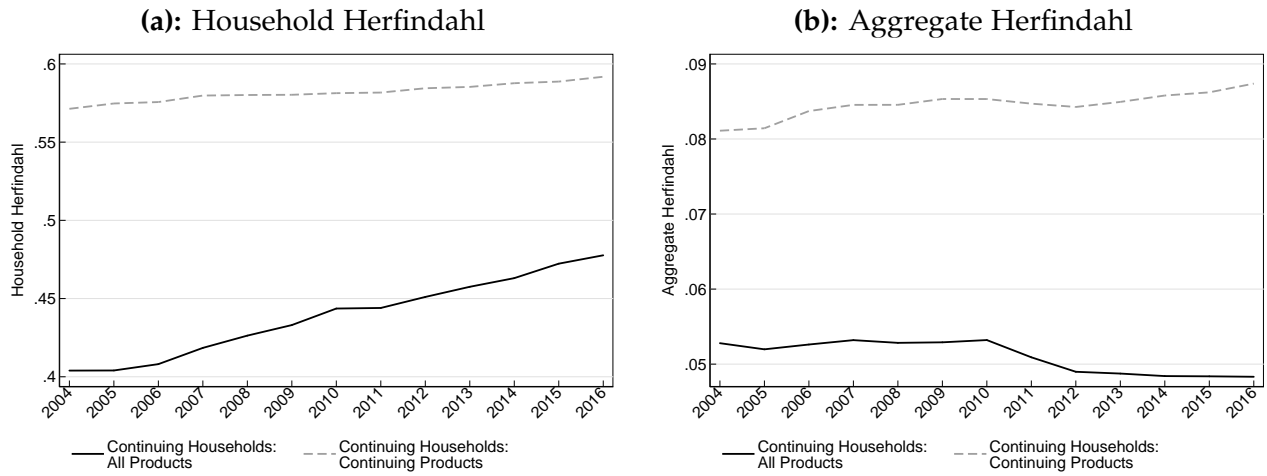


Table A1: Effect of Demographics on Household Concentration Trends

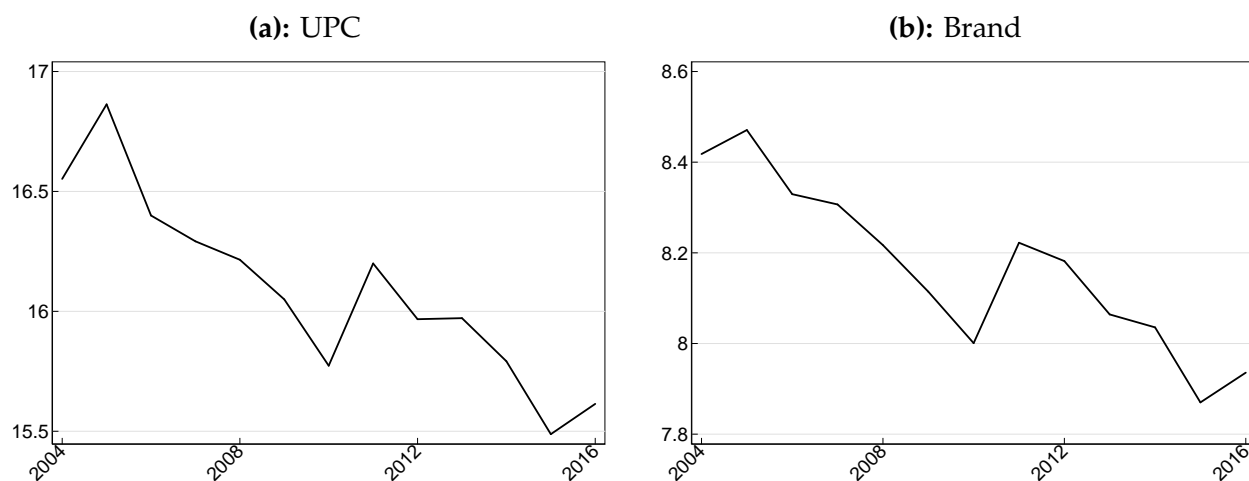
	(1)	(2)	(3)	(4)	(5)	(6)
	HH herf	HH herf	HH herf	HH herf	HH herf	HH herf
t	0.0026*** (0.00038)	0.0025*** (0.00037)	0.0024*** (0.00037)	0.0024*** (0.00037)	0.0034*** (0.00035)	0.0035*** (0.00035)
household size 2		-0.045*** (0.0029)	-0.044*** (0.0029)	-0.043*** (0.0031)	-0.032*** (0.0029)	-0.031*** (0.0028)
household size 3		-0.066*** (0.0050)	-0.062*** (0.0043)	-0.061*** (0.0046)	-0.048*** (0.0043)	-0.047*** (0.0043)
household size 4		-0.076*** (0.0062)	-0.070*** (0.0051)	-0.069*** (0.0054)	-0.051*** (0.0047)	-0.050*** (0.0046)
household size 5		-0.083*** (0.0068)	-0.077*** (0.0057)	-0.076*** (0.0061)	-0.054*** (0.0058)	-0.053*** (0.0057)
age 30-34			-0.0020 (0.0013)	-0.0018 (0.0012)	0.00066 (0.0013)	0.00034 (0.0013)
age 35-39			-0.0033* (0.0019)	-0.0032* (0.0019)	0.0030* (0.0016)	0.0031* (0.0016)
age 40-44			-0.0017 (0.0029)	-0.0016 (0.0030)	0.0093*** (0.0024)	0.0098*** (0.0023)
age 45-49			-0.000069 (0.0037)	0.000031 (0.0039)	0.014*** (0.0033)	0.015*** (0.0030)
age 50-54			0.0035 (0.0046)	0.0036 (0.0048)	0.018*** (0.0041)	0.019*** (0.0038)
age 55-59			0.0070 (0.0052)	0.0071 (0.0054)	0.023*** (0.0046)	0.024*** (0.0042)
age 60-64			0.0087 (0.0060)	0.0087 (0.0061)	0.027*** (0.0049)	0.028*** (0.0046)
age 65-69			0.0078 (0.0062)	0.0078 (0.0063)	0.025*** (0.0049)	0.026*** (0.0045)
age 70-74			0.017** (0.0070)	0.017** (0.0070)	0.031*** (0.0057)	0.031*** (0.0053)
Income \$20000-34999				-0.0021 (0.0023)	-0.00036 (0.0021)	-0.00078 (0.0018)
Income \$35000-59999				-0.0034 (0.0035)	0.0014 (0.0034)	0.000016 (0.0029)
Income >\$60000				-0.0026 (0.0052)	0.0071 (0.0052)	0.0024 (0.0041)
log spend					-0.065*** (0.0055)	-0.065*** (0.0055)
R-squared	0.31	0.32	0.32	0.32	0.38	0.38
N	49988358	49988358	49988358	49988358	49988358	49987989

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).

Appendix C. Measuring Varieties

Figure A15 shows the average number of UPCs purchased per household constructed using the same weights as were used in equations (3) and (5).⁵ Whether we define a product as a UPC or as a more aggregated brand, it is clear that the typical household is purchasing a smaller number of products across time.

Figure A15: Ave # Products Purchased by Each Household



It is also important to note that the typical household purchases many varieties within product categories. This motivates our modeling approach which includes love-of-variety effect at the *individual* level rather than the more standard macro model in which a representative agent has love-of-variety preferences which arise from aggregating over heterogeneous individuals who only consume a single product. We take this modeling approach because in our data context, individual households frequently purchase multiple products in narrow categories. This is true even if we focus on spending by single member households over short periods of time than our annual benchmark, so it is not driven solely by temporal aggregation or by multi-member households.

For example, focusing just on single person households, we find that for product-group weeks with positive spending, 40% have spending on 2+ UPCs, 18% have spending on 3+ UPCs, and 9% on 4+ UPCs. Again focusing on single person households but aggregating to monthly spending, we find 61% of product-group months have spending on 2+ UPCs, 37% on 3+ UPCs and 23% on 4+ UPCs.⁶

The variety statistics thus far focus on the number of products purchased by individual households.

⁵Since there is correlation across households in which products are consumed, the average number of products consumed by the typical household is not equal to the total number of products divided by the total number of households. For example, if all households consumed a single identical product, the average number of products per household would be 1 while the total number of products divided by the total number of households would be $1/(\# \text{ households})$

⁶These statistics weight individual households by their spending; these shares are reduced slightly if we weight households equally. We also still find frequent instances of purchasing multiple products if we define products as brands instead of UPCs but the shares are damped by around half.

We now turn to a discussion of *aggregate* variety availability and show that, due to both a statistical and conceptual complication, measuring the total number of products available (or purchased) in the economy is much more challenging. Thus, we treat aggregate variety availability as unobservable in our model. Importantly, in Homescan data, we observe only the set of UPCs which are purchased by households in the panel, not the set of all products which are purchased in the economy. While the Nielsen panel is large, a large number of products nevertheless are purchased by very few households and have tiny aggregate spending. The presence of a large number of products with very small sales means that in a statistical sense, it is very hard to measure entry and exit reliably due to sampling error. If we observe a product with no sales in period t-1 and very small total spending in period t, it is difficult to tell whether the product is newly available in period t, or if we just happened to not sample a household purchasing this product in period t-1. We can show *with certainty* that the Homescan panel does not capture the full set of products available in the economy, since we can observe products which have sales in the Nielsen retail data set but no sales in Homescan. For example, of all the UPCs which are ever purchased in a Retail Panel store, 25.5% are not purchased by a single household in Homescan. One might think that we could get around this by instead measuring products in the Retail Scanner data. However, this does not solve the problem, because this data is not a census of all stores. For example, we can see that 22.2% of UPCs which are purchased in Homescan are not sold in any store in the retail panel.

The more conceptual challenge, which would not be solved even if we had a full census of all US product sales, is that our model implies a distinction between products which are available and products which are purchased. We interpret products which are available but have no sales in our model as failed products. However, this is clearly an abstraction, and even the worst failed products will likely have tiny, but not actually zero sales. This means that even if there were no sampling issues related to products with small spending, we might still want to include some minimum aggregate spending threshold in order to “count” a product in the data.

Table A2 shows that the treatment of products with tiny spending in the Nielsen data indeed makes a huge difference for measures of aggregate varieties (both in levels and in growth rates), which is why we choose to treat this as an unobservable object in our model. For example, this table shows that although they represent only 2% of total spending, roughly half of all UPCs in the Homescan data have total annual spending across all households of less than \$25. Excluding products with small aggregate spending also leads to large changes in measured variety growth: counting products with even extremely tiny aggregate spending, delivers growth of 6.2% from 2004 to 2016, while dropping products with very tiny spending (which again are more sensitive to sampling error and interpretation issues) raises measured growth to 20-30%.⁷ Thus, while the data paints a robust pattern that the number of products is large and growing, exact product counts and growth rates are too uncertain to

⁷Note that for all of the calculations in Table A2, we compute statistics using a random subset of the Homescan Panel with a constant number of households per year so that statistics are not affected by changes in the panel size.

be usable as inputs to our model or resulting welfare inference.

Table A2: Effect of Products with Small Aggregate Spending on Statistics

Agg Spend Threshold	Share Spend > Threshold	UPCs per category 2004	UPCs per category 2016	UPCs per category % change	\mathcal{H}_{2004}^{HH}	\mathcal{H}_{2016}^{HH}	\mathcal{H}^{HH} % change	\mathcal{H}_{2004}^{Agg}	\mathcal{H}_{2016}^{Agg}	\mathcal{H}^{Agg} % change	Ω_{2004}	Ω_{2016}	Ω % change
\$0	100%	9248	9820	6%	0.262	0.284	8%	0.0036	0.0028	-21%	16.6	15.6	-6%
\$25	98%	4362	5193	19%	0.268	0.290	8%	0.0038	0.0030	-21%	15.8	15.0	-5%
\$50	95%	3153	3856	22%	0.275	0.296	7%	0.0039	0.0031	-22%	15.3	14.6	-5%
\$250	83%	1206	1555	29%	0.304	0.324	6%	0.0051	0.0039	-22%	13.2	12.7	-4%

In contrast, the statistics which are the focus of our analysis (household and aggregate Herfindahls as well as the average number of products purchased by individual households) are very robust to the treatment of these products with small aggregate spending, since these statistics depend much more on products with substantial spending. Growth rates of these variables (which are more important for our model inference) are even more stable across these spending thresholds, changing by at most a couple percentage points when moving from no aggregate spending threshold to a fairly restrictive threshold. An advantage of our modeling framework is thus that we can infer product availability changes and their welfare consequences using these observable statistics even though we cannot reliably measure product availability itself.

Most importantly Table A2 shows that our model inference for variety availability and welfare are almost completely unaffected by the behavior of these products with small aggregate spending. Performing inference on statistics constructed using spending on all products produces nearly identical conclusions to inference performed on statistics which exclude products with tiny or modest aggregate spending. For example, in all cases, the model implies that the increase in consumed varieties (j^*) is always 27-28%, and that annual welfare growth when fully accounting for all of the time-trends in the data is 0.45%-0.47%. Thus, none of our model conclusions are affected by the behavior of the large set

Table A3: Effect of Products with Small Aggregate Spending on Model Results

Agg Spend Threshold	j^* growth	\tilde{N} growth	η growth	Utility growth from N	Utility growth from N, F, σ
\$0	27%	70%	-2.20%	0.56%	0.45%
\$25	27%	70%	-2.5%	0.57%	0.46%
\$50	28%	71%	-2.6%	0.57%	0.47%
\$250	28%	70%	-3.2%	0.56%	0.47%

of UPCs in Nielsen with negligible spending.

Appendix D. Relationship to External Data

D.1 External Spending Data

Figure A2 shows that aggregate Nielsen spending lines up well with spending growth measures from the Consumer Expenditure Survey and BEA national accounts for similar categories.⁸

However within-household spending growth is substantially less strong than overall household spending. This is likely driven by two forces: 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., they do not provide weights to make the panel dimension 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 fully balanced and within-household spending growth patterns. 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 A8 thus redoes Figure 1 using a fully balanced panel and with a specification using only the within-household changes specification.

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

⁸It is well-known that the consumer expenditure captures a lower level of spending than the BEA and this "missing spending" has a positive trend. However this growth in missing spending mostly occurs prior to our sample period. Throughout our sample period, the CEX captures a relatively constant share of aggregate spending. This means CEX spending growth is slightly lower but broadly similar to aggregate spending growth from the BEA.

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.

D.2 Census Concentration of Production

A large and growing literature uses production data from the Census to show that the concentration of production has been broadly increasing from 1982-2012. For example, [Autor](#), [Dorn](#), [Katz](#), [Patterson](#),

and Reenen (2017) calculates industry concentration within 4-digit industries, and averages this up to 6 major sectors and shows that various concentration measures have all increased when comparing 1982 to 2012. In this section we explore the relationship between the concentration measures in our paper and this large literature and argue that relevant comparisons from Nielsen data are broadly consistent with this Census based literature.

First, it is important to note that the concentration notions we emphasize in our paper are conceptually distinct along a number of important dimensions from the concentration of firms or establishments studied using census data. Most importantly, we are measuring the concentration of spending over very detailed UPCs (or slightly coarser but still highly disaggregated brands). This is a fundamentally much more disaggregated notion of concentration than that studied with production data, since firms can potentially produce tens, hundreds or even thousands of different products. For example, in our data Procter and Gamble produces over 40,000 unique UPCs, L'Oreal produces over 28,000 UPCs and General Mills, Unilever, and Kraft Heinz all produce 10,000-20,000 UPCs.⁹

Furthermore, the categories within which we calculate concentration are also more disaggregated than those in typical Census-based calculations and also cover a more narrow subset of production. For example, the broad manufacturing sector in Autor, Dorn, Katz, Patterson, and Reenen (2017) covers 86 4-digit industries within which concentration is computed. However, of these 86 industries only a small subset produce in categories which are covered by Nielsen (for example NAICS Code 3111 "Animal Food Manufacturing") while most are in production industries which have no overlap with Nielsen categories (for example NAICS Code 3336 "Engine, turbine, and power transmission equipment manufacturing" or NAICS Code 3365 "Railroad rolling stock manufacturing").

Finally, it is important to note that our sample covers the period 2004-2016 while census data starts in 1982 and is last available in 2012. The exact timing of concentration trends in Census data varies substantially, with many sectors exhibiting increases primarily in the period prior to our sample period.

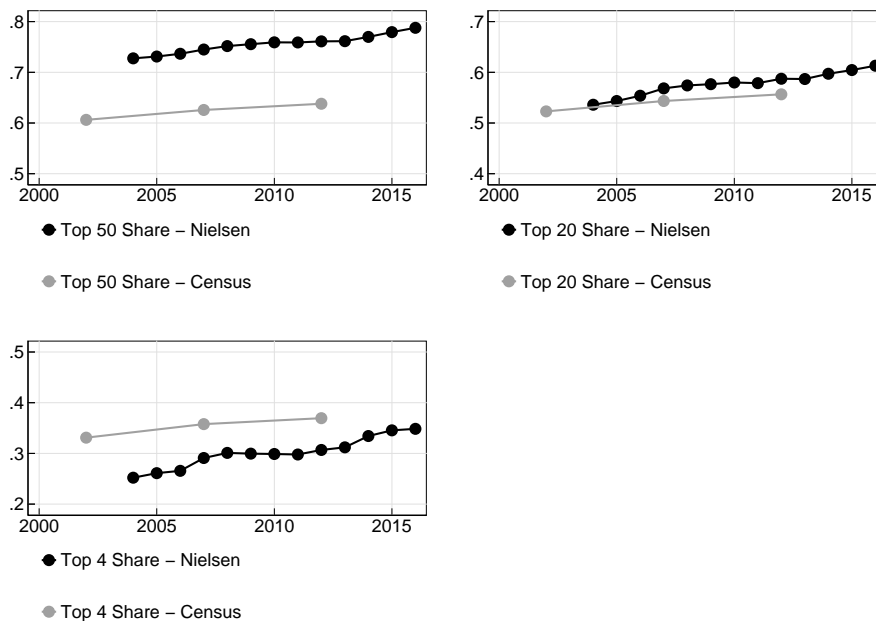
Since they are conceptually different notions, this means the aggregate product concentration trends which we emphasize in the body of the paper should not be directly compared to production concentration trends in Census. However, we can construct concentration measures using the Nielsen data which *are* more comparable with Census calculations and that can be used to explore the external validity of our data. We now explore these comparisons.

Since households in the Nielsen sample report the retail chain in which they shop, we can aggregate up total spending to compute a Nielsen based measure of spending at each retail chain and resulting retailer concentration. This can then be compared to the concentration of retail trade in Census data. Specifically, since the Nielsen sample is focused on grocery and drug store spending, in the Census we use firm concentration numbers only from NAICS Code 445 "Food and beverage stores" and 446

⁹It is also worth noting that our "household" concentration measures have no analogue in the Census literature even if we were measuring producer rather than product concentration.

"Health and personal care stores" and weight the publicly available Census concentration numbers for these two sectors using their relative share of sales. This clearly does not provide a precise match between the retail establishments covered in Nielsen and Census so we should not expect numbers to line up exactly, but Figure A16 shows that that Nielsen data broadly matches the level of retail spending accounted for by the Top 4, Top 20 and Top 50 firms as well as the upward trend in retail concentration.

Figure A16: Retail Trade Concentration



We can also perform a similar exercise by allocating UPC-level spending up to the manufacturer. When manufacturers produce a new product, it is assigned a barcode by the company GS1, which then maintains a database which can be used to link UPCs to manufacturers. This lets us aggregate product spending up to a measure of manufacturer spending, with two important caveats:

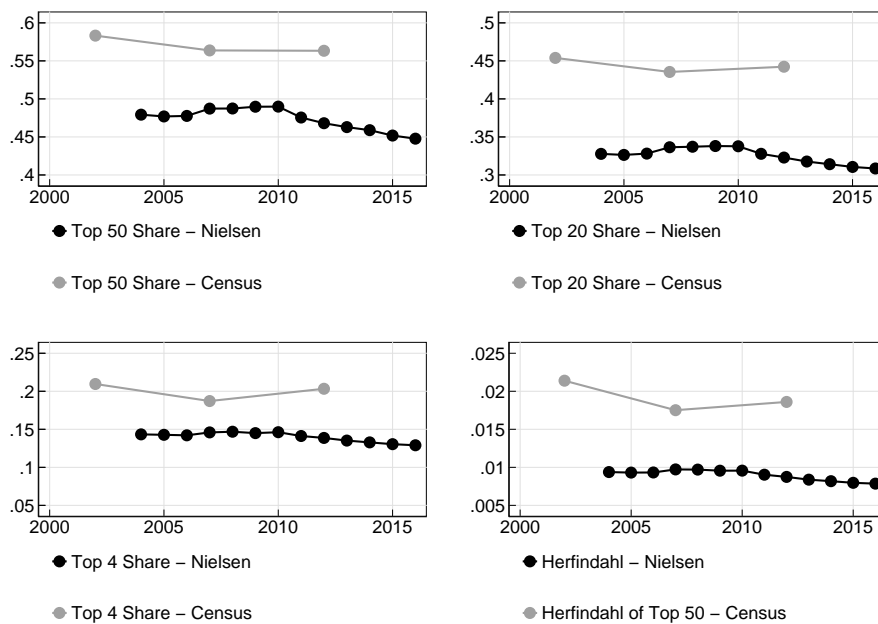
First, the link from UPCs to parent companies is sometimes inconsistent. For example, Gillette and Old Spice were both acquired in the past by Proctor and Gamble, and the UPCs for Gillette and Old Spice products both map to Proctor and Gamble. However, Ben and Jerry's was acquired by Unilever in 2000, yet UPCs for these products are assigned to the "Ben and Jerry's Homemade Inc" firm name rather than to the Unilever parent company. Similarly, Goose Island Beer UPCs are assigned to "Goose Island Beer Company" even though this firm was acquired by InBev in 2011. To the extent that some UPCs are assigned to subsidiaries rather than parent companies, our Nielsen based measure of manufacturer concentration will be biased downwards.

Second, UPCs for store-brand products map to the retailer rather than the actual manufacturer of

the product. For example, Costco's "Kirkland" store-brand barcodes all map to "Costco", even though Costco does not actually produce most of these products. Although sometimes the actual producer can be identified (for example Kirkland Coffees are advertised as being roasted by Starbucks), this information is typically a trade-secret. This means that we cannot measure the producer for most generic products, and as a result we must drop these products when aggregating up UPCs to manufacturers and focus only on branded products. To the extent that the production of generic products is proportional to the production of branded products, this will have no effect on concentration. However, it is likely that generic products are disproportionately produced by larger manufacturers, so dropping generic products is likely a second force that will bias our Nielsen based measures of manufacturer concentration downwards.

To again focus the comparisons on the most relevant producers, we keep NAICS codes 311 and 312 "Food Manufacturing" and "Beverage and Tobacco Product Manufacturing" from the Census data and weight these concentration measures by their relative sales shares. Figure A17 shows that despite the above concerns, Nielsen data again broadly matches Census data, producing similar levels of manufacturer concentration and a flat to mild downward trend.

Figure A17: Manufacturer Concentration



Overall the results in these two subsections give us confidence that the Nielsen data is largely in line with external evidence on aggregate spending and with Census data on producer concentration.

Appendix E. Model Simulation Results

In this section, we explore numerical simulations of our model to test the validity of our elasticity approximation as well as to explore how restrictive the assumption of a stable distribution of Pareto taste-adjusted prices is for our conclusion.

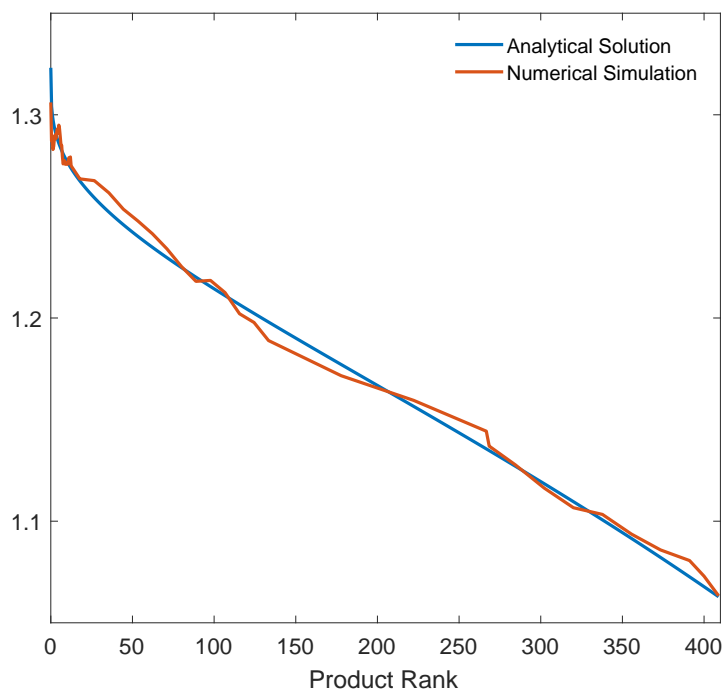
We simulate a discrete approximation to the main model in the paper by drawing a large random vector $\tilde{\gamma}^{rand}$ of price-adjusted tastes from a Pareto distribution for a large sample of households, using the same parameters as our baseline model. While $N = 15000$, our baseline random sample uses 2.25 million draws for each of 20,000 households since we are trying to approximate a continuum of products from $[0, N]$. However, rather than using analytical formulas to calculate household market shares, for each household we instead keep the random set $\Omega = \tilde{\gamma}^{rand} > \tilde{\gamma}^*$ and then calculate a numerical price index directly from $P = \left(\int_{k \in \Omega_i} (\tilde{\gamma}_{i,k})^{\sigma-1} dk \right)^{\frac{1}{1-\sigma}}$ and then compute market shares from Equation 11. These formulas hold for arbitrary distributions of taste, so even though we still simulate the taste draws from a Pareto distribution, in this simulation we are using no analytical results that rely on this assumption, which also means that we can also perform a similar procedure even if tastes do not follow a Pareto distribution.

In order to get aggregate market shares, we must identify the particular products that each household consumes. In order to do so, we use our rank function Equation 13 with a random uniform draw to compute for each household, the aggregate ranking of each of the 2.25 million possible products in $[0, N]$ and then compute household i 's particular idiosyncratic rank for each of the 2.25 million j products. We then sort $\tilde{\gamma}^{rand}$ and map the highest value to $r_{ij} = 0$, the second highest value to $r_{ij} = 1$ and so on to $r_{ij} = 2.25$ million. Finally, since for each r_{ij} , we know the value of j , this means that we then know household i 's taste draw and resulting individual spending for each aggregate product j . For example, the households highest $\tilde{\gamma}^{rand}$ draw will always map to their $r_{ij} = 0$, but the corresponding aggregate j which household 1 ranks highest might be $j = 0$, the j which household 2 ranks highest might be $j = 2043$, and the j which household 3 ranks highest might be $j = 17$. Once we have these household specific spending shares for each product j , we can then numerically add up total spending on each product j to calculate aggregate market shares.

Since these are computed entirely numerically, they do not rely on any of our closed form solutions for aggregate market shares and are thus again valid even under departures from the Pareto distribution. As we note in 4.5, our analytical market shares are only valid under the Pareto distribution so we must approximate the elasticity of demand by modeling a price change as a switch with another product in the aggregate ranking. Since these numerical results do not rely on the Pareto distribution, we can use this numerical model to simulate the aggregate elasticity of demand and resulting markup for a product j by just raising all households' random taste draw for that product by a small amount. Note that calculating elasticities for each j requires re-simulating a new set of aggregate market shares.

For these sample sizes, computing an elasticity for a single j requires roughly 2 hours of computational time, so it is infeasible to simulate the elasticity of demand for all 2.5 million products. Instead, we compute the elasticity of demand and implied markups for 50 different values of j distributed throughout the product space. Figure A18 compares this simulated markup to our analytical approximation and shows that the analytical approach produces essentially identical results (noting that there is still obvious numerical simulation error even with these large sample sizes).

Figure A18: Simulated vs. Analytical Approximation for Markup



As stressed throughout the paper, our analytical derivations and implications of changes in N are only valid under the assumption that the distribution of price-adjusted tastes continues to follow a Pareto distribution as we vary N . If markups were fixed for all products, then assuming that the distribution of price-adjusted tastes is held fixed as N varies would be a natural benchmark. However, our model instead implies that optimal markups do vary across products, and that the markups for individual products change as we vary N . This implies that if household tastes for products and their marginal costs held fixed, but we allow prices to change along with optimal markups when N changes, then there will necessarily be a violation of the assumed Pareto distribution. Since all of our analytical results assume the Pareto distribution of price-adjusted tastes, this means that our analytical comparative statics to changes in N and F which induce changes in product markups are technically comparative statics in response to these parameter changes plus whatever implicit changes in household tastes (or marginal costs) are necessary to preserve a Pareto distribution of price-adjusted tastes after markups adjust. In practice, high turnover means that the set of products purchased in 2004 and

in 2016 is mostly disjoint, so one can primarily interpret these as taste shifts for new products rather than taste changes for existing products.¹⁰ However, if the required taste shifts necessary to maintain the Pareto distribution under our counterfactuals were substantial, then this would potentially substantively change the interpretation of the welfare effects of changes in N .

However, we now use our numerical model to show that even though there are indeed implicit taste changes necessary to maintain the Pareto distribution as N changes, in practice these required taste changes are quantitatively small and actually work against our conclusion that N is welfare improving. We thus conclude that even though this is a large potential issue for the interpretation of our comparative statics, it is of little quantitative importance in practice. Specifically, we perform the following exercise: For the initial value of N in 2004, we simulate our numerical model exactly as described above. Given household i 's resulting distribution of tastes for all j products $\tilde{\gamma}_{i,j}^{rand}$, we can then compute a household's actual (non-price adjusted) taste for product j $\gamma_{i,j}^{rand} = \tilde{\gamma}_{i,j}^{rand} \mu_j$ using the analytical formula for μ_j from Section 4.5.¹¹ Note that as we explore above, even though our numerical model does not otherwise rely on analytical results, this analytical formula for the markup is valid since we are drawing the numerical distribution of price-adjusted tastes in the model from a Pareto distribution.

We then increase N in the model but hold the particular random realizations of $\tilde{\gamma}_{i,j}^{rand}$ exactly fixed in the new simulation. Thus, by assumption, the values of *price-adjusted* tastes will be identical in the two simulations. However, as N increases, the function μ_j and resulting prices will change. If price-adjusted tastes are fixed by assumption, but prices change then household tastes must change.¹²

How large are the required taste changes necessary to maintain an identical realization from a Pareto distribution of price-adjusted tastes as N increases? Figure A19 shows that these changes are small. The left panel plots the implied taste draws as a function of initial aggregate product rank j for a fixed household before and after a 70% increase in N .¹³ Clearly the increase in N induces some implied changes in tastes in order to maintain the Pareto distribution for price-adjusted tastes, but it is also clear that the requisite taste changes are small. The right panel of the plot shows a scatter plot of the realizations of taste before and after the increase in N . Overall the R^2 is above 0.999, so there is an almost perfect correlation of tastes under the two scenarios. In order to maintain an identical distribution of price-adjusted tastes, there is a modest *decline* in the implied average taste when N increases, which lowers implied welfare by roughly 1.3%. This occurs because as N increases, markups for incumbent products decline, which makes price fall and thus taste/price rise. In order to maintain a constant taste/price for that product, this means taste for those products must decline.

¹⁰Only 13.2% of UPCs purchased in 2004 are still purchased in 2016.

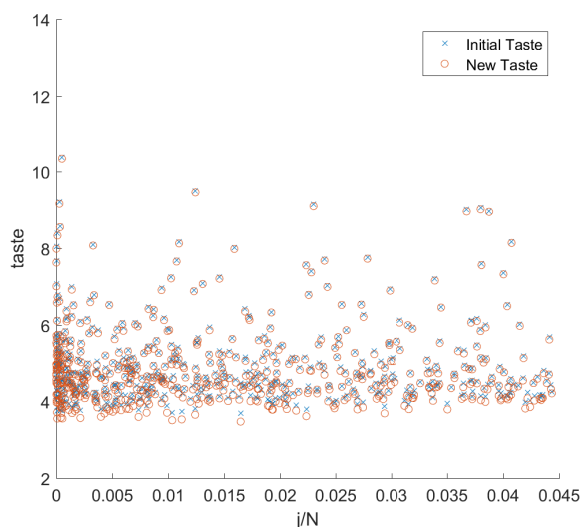
¹¹For notational simplicity, we assume that marginal cost is 1 for all products. More generally this approach actually recovers the distribution of marginal cost adjusted tastes. As long as we assume marginal cost is constant as we vary N , one can interpret changes in taste and changes in marginal cost adjusted taste equivalently so these are equivalent exercises.

¹²Further, since markup changes are a monotonic function of j but individual rankings of the j products are non-monotonic when $\alpha > 1$, these price changes will be non-monotonic over individual households' consumption baskets.

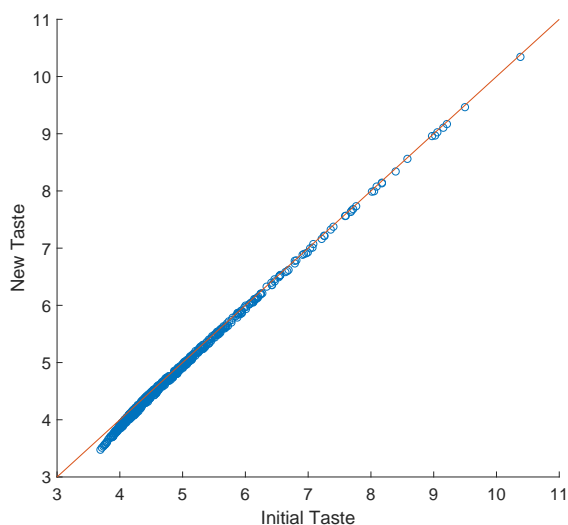
¹³Here we focus on products which are consumed in both scenarios so that such taste comparisons are relevant.

Figure A19: Household Taste Changes Required to Maintain Pareto with N increase

(a): Initial and New Tastes by Aggregate Product Rank



(b): Initial vs. New Tastes



However, the welfare conclusion in the body of the paper under the assumed constant Pareto distribution of price-adjusted tastes is that an increase in N of 70% raises welfare by roughly 9.5%. The numerical results above show that those welfare results are only valid if there is also a simultaneous modest decline in non-price adjusted tastes when N rises, suggesting that if one instead held tastes fixed when increasing N and departed from Pareto, the welfare increase would be slightly stronger. While such an exercise could potentially be performed numerically, it would require solving for the entire equilibrium distribution of the elasticity of demand and resulting markups numerically. As discussed above, the numerical calculation of the elasticity of demand (even for a single product in partial equilibrium) is very computationally costly.

Finally, we use this simulated model to also explore the role of potential measurement error in driving concentration trends. Although [D](#) shows that the Nielsen data tracks aggregate spending measures fairly closely, the declining within-household spending patterns suggest there may be some role for attrition related measurement error across time. Furthermore, even though households are supposed to report online purchases and that [Figure A1](#) shows that online spending is relatively unimportant for these sectors, it is possible that under-reported online spending might also drive increasing measurement error across time.

While it is difficult to analytically characterize the role of various forms of measurement error for concentration trends, we follow the indirect inference approach in [Berger and Vavra \(2015\)](#) and [Berger and Vavra \(2019\)](#) and *simulate* various flexible forms of measurement error in the numerical version of our model under the assumption that all other model parameters are held fixed. Specifically, we simulate the discrete version of our model and separately consider the effects of measurement error

on household and aggregate concentration. We focus primarily on measurement error arising from failing to report transactions entirely rather than from misreporting the size of a transaction, since the former is much more likely given the structure of the Homescan data collection. We consider three types of potential under-reporting encompassing various different extremes: 1) households failing to report some randomly chosen purchases, 2) households failing to report their smallest purchases and 3) households failing to report their largest transactions. Overall, we find that while measurement error can change both household and aggregate concentration, it pushes both household and aggregate concentration in the same direction and so is unlikely to be an important explanation for the observed rise in niche consumption. Unsurprisingly, the first and second form of measurement error raise both household and aggregate concentration while the third form of measurement error instead lowers both concentration measures. Furthermore, the second form of measurement error seems most plausible given the nature of the Nielsen data, since a household might fail to report a small one-off purchase which is likely to be a small share of that household's annual spending but is unlikely to consistently fail to report large, regular purchases that are likely to be a large share of annual spending. Since the third form of measurement error is especially unlikely, this means that measurement error is then also quite unlikely to explain a decline in aggregate concentration. As emphasized in Section ??, a decline in aggregate concentration with flat household concentration would generally be sufficient to infer an increase in N . Overall, these simulation results strongly suggest that measurement error does not drive the rise of niche consumption.