The Rise of Niche Consumption*

Brent Neiman
University of Chicago

Joseph Vavra
University of Chicago

February 2021

Abstract

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 purchase different products from each other. As a result, aggregate spending concentration has decreased. We develop a model of heterogeneous household demand and use it to conclude that increasing product variety drives these divergent trends. When more products are available, households select products better matched to their tastes. This delivers welfare gains from selection equal to about half a percent per year in the categories covered by our data. Our model features heterogeneous markups because producers of popular products care more about their existing customers while producers of less popular niche products care more about generating new customers. Surprisingly, our model matches the observed trends in household and aggregate concentration without any 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 we thank our discussants Jorge Miranda-Pinto and Mingzhi Xu. We also thank 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 over the last 15 years, the typical household has dedicated an increasing share of its expenditures to a few preferred products. At the same time, households are increasingly buying different products from each other. Thus, aggregate spending concentration has declined even as household spending concentration has risen. We develop a novel model of heterogeneous household demand to interpret these new facts and explore their implications. In the model, as new products are introduced, households can choose consumption bundles better suited to their particular tastes, resulting in welfare gains from better selection. We fit our model to the data for more than one hundred separate product categories and conclude that rapid growth in product availability has led to sizeable welfare gains that cannot be identified with standard representative agent macro models.

We begin our analysis by using the Nielsen Homescan dataset that covers a large fraction of spending on groceries and other household nondurables to study the shopping behavior of thousands of households from 2004-2016. We measure household-specific product spending shares within narrow categories like "Coffee" and "Cosmetics" and demonstrate that these shares have steadily become more concentrated over time. 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. These diverging household and aggregate concentration trends imply that even though each household increasingly focuses spending on its own preferred products, households also increasingly differ in 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 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 rich and poor households, between consumers in one region and another, or by differences between households grouped according to various 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? As discussed below, these trends in concentration are of similar magnitudes to firm-level concentration trends which have received significant attention in recent years. Nonetheless, it is difficult to assess the importance of these trends without a model that links them to underlying fundamentals. Standard models cannot be used for this purpose 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 build a model of consumers that have a love-of-variety and whose preference orderings for particular products differ from each other.

Our model of individual household demand follows Li (2019). It features constant elasticity of substitution (CES) preferences, product-specific tastes, and a utility cost borne per variety consumed. Households take the distribution of prices as given and typically choose to consume only a subset of the total available products. When taste-adjusted prices are distributed Pareto, we obtain a closed-form expression relating the Household Herfindahl index – the concentration measure we use in our empirical analyses – to structural parameters of the model. In Li (2019), however, households all have identical tastes and agree on the ranking of products, so there is no capacity for the divergent household and aggregate trends that we find in our empirical results.

The key innovation in our model comes from our novel and analytically tractable specification of household heterogeneity, which is crucial to jointly match household and aggregate product demand patterns and to speak to the rise in niche consumption in the data. We introduce a continuum of households with correlated but heterogeneous preferences for different products in a way that yields closed-form solutions for both household and aggregate spending concentration. In particular, we assume that all households have tastes that decline identically from their favorite product to their second favorite, and so on, so that all households have identical taste *distributions*. However, the actual identities of these first- and second-favorite products are allowed to 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's price is reduced, sales to existing customers expand with a constant elasticity, but the product

[1] Dynamic 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, as we shown in Appendix A.3.

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 analytically the distribution of elasticities as well as the implied aggregate markup in the economy.

Our closed-form analytical solutions for various aggregates rely on particular parametric assumptions, but we show that, nonetheless, the model matches product-level spending distributions both for individual households and for the economy as a whole quite well. It is important in our setting to have analytical expressions for observed empirical moments 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 we demonstrate in Appendix A.3, the level and trend growth in the total number of products are highly sensitive to the choice to include or exclude products accounting for tiny amounts of aggregate spending. 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, a rate of roughly 4.5 percent per year. This increase in available varieties leads households to endogenously consume more products and enjoy welfare "gains from variety", 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 to 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 4.5 percent annual growth in the number of available products into our model, we calculate that consumption-equivalent welfare grows by 0.56 percent per year for expenditures in the categories covered by our data, and that 93 percent of this growth comes from gains from selection. 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 match 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 multiple configurations are consistent with the data.[2] However, we consider various combinations of parameter changes in addition to the increase in available 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 pieces of empirical evidence that support this relationship.[3]

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 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 with losses 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 have more limited effects on welfare.

In the final part of the paper, we explore the implications of the rise of niche consumption for aggregate 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 inferred changes in structural parameters that drove the rise of niche consumption? Sur-

[2] For 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.

[3] For 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.

prisingly, not much. In particular, as we demonstrate analytically, increases in product availability do not on their own 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.

By contrast, changes in the elasticity of substitution and in the shape of the taste distribution have the scope to affect aggregate market power, but the changes suggested by our model are not large enough quantitatively to meaningfully alter the picture. Our model demonstrates that the significant trends in household and aggregate product concentration need not indicate any changes in aggregate market power.[4] Even more broadly, our model demonstrates how any given trend in concentration can be associated with an increase or a decrease in aggregate market power depending on the underlying structural forces driving the trend.

We proceed as follows. Section 2 discusses the related literature, Section 3 demonstrates the 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 (2019). Our theoretical contribution is to maintain analytical tractability even when extending this setup much further to an environment with heterogeneity in which households have different but potentially correlated tastes across products. This allows us to speak to the increasing divergence between household and aggregate concentration. Our basic approach follows in the tradition of the macro and trade literature that uses a CES structure to study the implications of expanded product availability.[5] The heterogeneous and asymmetric preferences in our model imply that expanding the set of available products benefits consumers through a selection effect that is above-and-beyond the standard love-of-variety gains in symmetric representative agent models.[6]

[4]Similar conclusions also obtain when we compute aggregate market power trends separately for each of the product groups, although there are a few exceptions such as photographic supplies.

[5]For example, Handbury and Weinstein (2014) emphasizes the need to account for differences in variety availability when comparing the price level across U.S. cities. Redding and Weinstein (2016) demonstrates how welfare measures can be biased if they do not account for heterogeneity in consumer tastes across products. Atkin *et al.* (2018) uses similar scanner data on grocery purchases to calculate the welfare gains associated with entry of global retail chains into the Mexican market.

[6]While representative agent CES models can be rationalized through an underlying discrete choice representation with heterogeneity (Anderson *et al.* (1987)), this requires idiosyncratic tastes to be drawn from an i.i.d. Gumbell distribution and aggregates to an environment with symmetry.

Our analysis also relates to a large literature in industrial organization (IO) quantifying the welfare gains from new varieties.[7] 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). 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 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 *et al.* (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 *et al.* (2018a,b) shows 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 *et al.* (2017) and Furman and Orszag (2015). Our finding that household and aggregate concentration trends move in opposite directions is reminiscent of Rossi-Hansberg *et al.* (2018), which demonstrates 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 *et al.* (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 *et al.* (2018), Stroebel and Vavra (2019), Anderson *et al.* (2018), and Burstein *et al.* (2019). We note, however, that our model

[7]See, for example, Hausman (1996), Petrin (2002), and Brynjolfsson *et al.* (2003).

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 *et al.* (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 *et al.* (2015); Alwin and Tufis (2015)), in political idiology (Pew Research Center (2014); Gentzkow *et al.* (2017)), in jobs and income (Autor *et al.* (2006); Piketty *et al.* (2016)), and in the geography of where households consume (Davis *et al.* (2017)).[8] Our findings indicate that, along with these other manifestations of fragmentation in modern life, even our grocery purchases increasingly differ from the national average.

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.1. 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.[9] 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. We focus on the full sample of households for a number of reasons discussed in Appendix A.2, but this is relatively

[8] The 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.

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

conservative since the magnitudes of our trends increase when we restrict to a balanced panel of households.[10]

In our baseline analysis, we define a product as a Universal Product Code (UPC). Appendix Figure A10 demonstrates, however, that the key trends we identify are robust to instead defining a product as a "brand". Nielsen assigns 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.[11] Sales of generics are large and growing, so their inclusion, by construction, distorts concentration measures that define products as brands.[12] 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 also note that our analysis focuses on product level concentration treats all products symmetrically within a category and does not treat differently products produced by the same manufacturer from those produced by different manufacturers. This is primarily because data limitations discussed in Appendix A.4 make it infeasible to accurately measure products produced by the same manufacturer. However, even if data was available to group UPCs produced by the same manufacturer together, it is not obvious that new niche product varieties which cater to particular idiosyncratic tastes should enter differently in household preferences depending on the nature of the production process. That is, it is not clear that grouping multi-product firms together would be a preferable specification, even if it was empirically feasible.

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).

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

[11]Since we do observe UPC codes just not brand labels for generic products, a generic product at one retailer will count as distinct from a similar generic product at a different retailer in the typical situation in which they have different UPCs.

[12]See, for example, Dube *et al.* (2018). Our robustness checks using the brand definition of product exclude generics.

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:

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

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:[13]

$$\mathcal{H}^{\mathrm{HH}}_{i,c,t} = \sum_j \left(s_{i,c,j,t} \right)^2. \tag{2}$$

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

$$\mathcal{H}^{\mathrm{HH}}_{c,t} = \sum_i share_{i,c,t} \mathcal{H}^{\mathrm{HH}}_{i,c,t}, \tag{3}$$

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

$$share_{i,c,t} = \frac{\sum_j \left(\omega_{i,t} E_{i,j,c,t} \right)}{\sum_i \sum_j \left(\omega_{i,t} E_{i,j,c,t} \right)}, \tag{4}$$

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:

$$\mathcal{H}^{\mathrm{HH}}_{t} = \sum_c share_c \mathcal{H}^{\mathrm{HH}}_{c,t}, \tag{5}$$

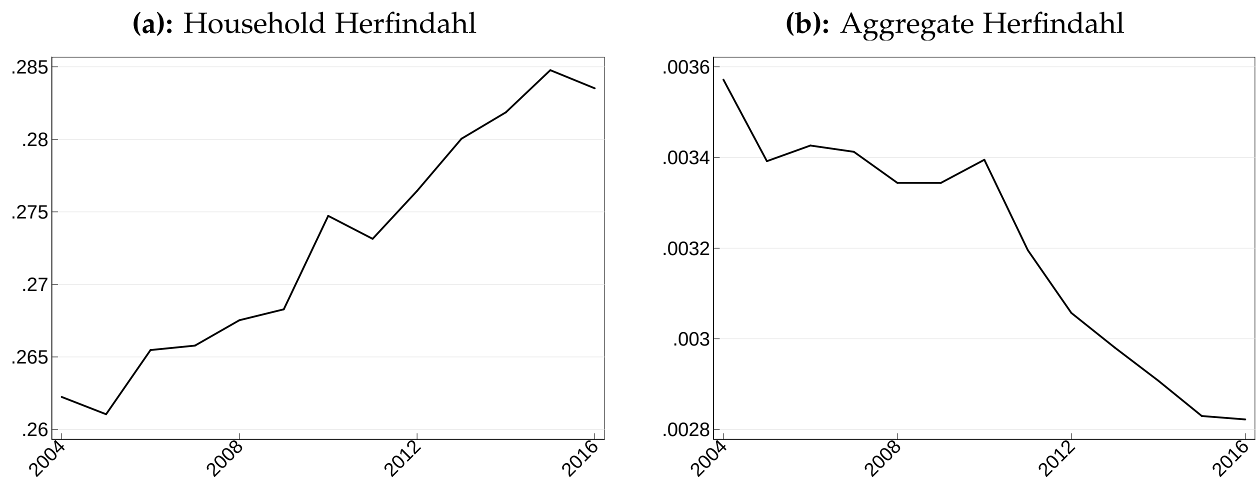
where $share_c$ is the average share of category c in total spending across our entire sample.

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 those emerging from shifts in spending across categories with different average 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. However, our results are robust to instead allowing compositional shifts across categories to influence our concentration measures.

Figure 1a plots $\mathcal{H}^{\mathrm{HH}}_{t}$ and reveals a nearly monotonic increase in household spending concentration

[13] All 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.

Figure 1: Household Product Concentration



from 2004-2016. In Appendix A.3, 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 this series with a linear trend yields a precise and highly significant estimate.

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 *et al.* (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:

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

and the Aggregate Herfindahl in category c as:

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

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:

$$\mathcal{H}_t^{\text{Agg}} = \sum_c share_c \mathcal{H}_{c,t}^{\text{Agg}}. \tag{8}$$

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.

Are these trends big or small? One simple way of contextualizing their scale is to compare to the trends in firm concentration that have received significant attention in the recent literature including Autor *et al.* (2017). In Census data the Herfindahl of firm sales in the median 3-digit manufacturing sector grew by 0.5 percent annually from 2002-2012. By comparison, we find annual household-product level Herfindahl growth of 0.65 percent and aggregate product-level Herfindahl growth of -1.94 percent in our data.

The decline in aggregate concentration we find might, at first, seem at odds with the rise in sales concentration measured in Census data. Our aggregate concentration measure, however, captures expenditures in a different set of sectors, spans a different time period, and aggregates to the product level whereas Census-based estimates aggregate products up to the producer level. The resulting trends may therefore differ significantly. Appendix A.4 relates our findings to this recent literature on rising market concentration in the U.S. census data. 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, 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. 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 A.3, 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 Robustness to Measurement and Composition

The rise in niche consumption – the increase in the Household Herfindahl and decrease in the Aggregate Herfindahl – is highly robust and is not driven by either measurement choices or by obvious composition effects. Appendix Figures A8-A14 show that these divergent concentration 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, focus on a balanced panel of households over time, or condition on household size. While the increase in household concentration and decline in aggregate concentration holds across all specifications, the exact levels of concentration and quantitative magnitude of trends unsurprisingly varies somewhat with measurement choices.[14] However, targeting the theoretical model we develop below to match these alternative empirical results generates very similar conclusions, so in that sense our conclusions are robust to all of these measurement choices.

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. To show this, we recalculate 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. The rising Household and falling Aggregate Herfindahls do not simply owe to changes in composition across groups with different levels of concentration. Appendix Figure A13 further shows that we also see a rise in niche consumption when restricting to a balanced sample, so changes in the composition of the household sample over time do not drive our results.[15] Our results are also similar if we foresake Nielsen weights and instead weight households equally.

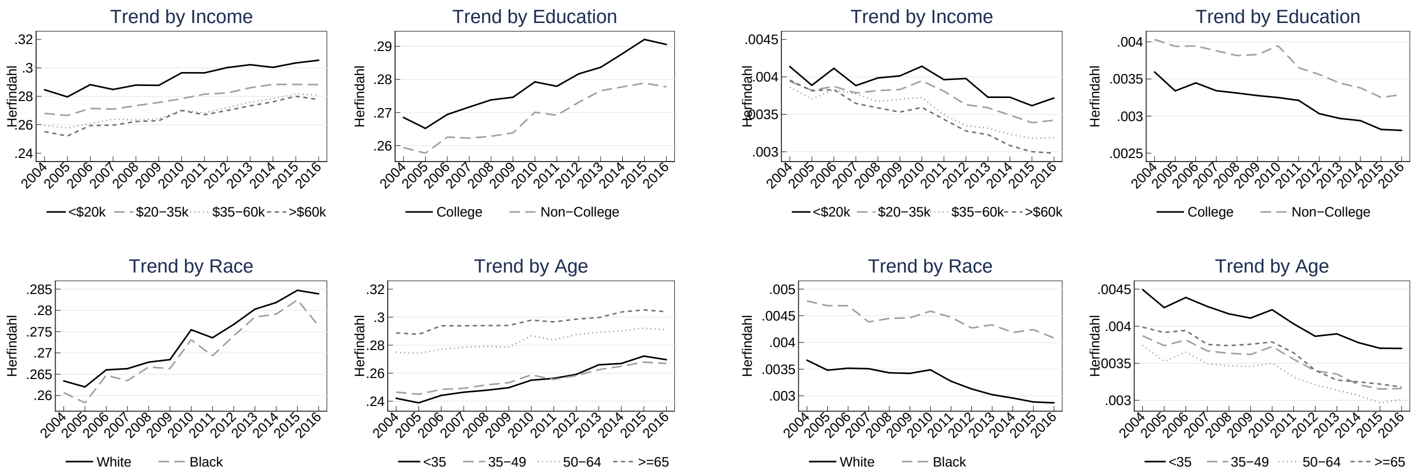
As a simple summary statistic for the prominence of niche consumption, we consider the ratio of the Household Herfindahl to the Aggregate Herfindahl. A higher value for this "niche ratio" means that household consumption is more segmented into different niches. Figure 3 shows that the rise of

[14] Aggregate brand concentration exhibits more volatility but an overall decline from 2004-2016. However, we again refer to the discussion in 3.1 about how Nielsen's method for brand assignment does not allow for a consistent measurement of concentration. Brand-level evidence should be interpreted with caution relative to our main UPC specification.

[15] Trends are actually stronger in the balanced panel specification. See Appendix A.2 for discussion and interpretation.

Figure 2: Trends within demographic group

(a): Household Concentration **(b):** Aggregate Concentration



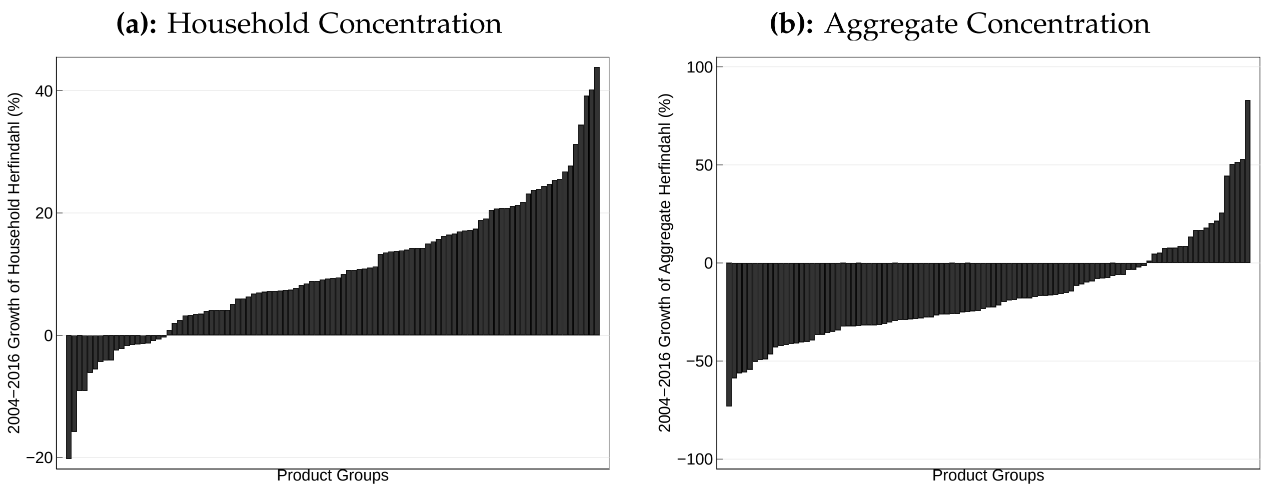
niche consumption is pervasive across product categories, with three-quarters of product categories exhibiting increases in $\mathcal{H}_{c,t}^{\text{HH}}$, 80 percent of product categories exhibiting decreases in $\mathcal{H}_{c,t}^{\text{Agg}}$, and growth in the niche ratio in over 90 percent of the categories.[16]

In Appendix Figure A15, 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 or differences across 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. Appendix Figure A16 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 observe are not simply driven by shifts in where households shop. Appendix Figure A17 shows the patterns hold even *within* retailers so that patterns are not driven by compositional changes in which retailers households shop at.

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".

[16]To 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.

Figure 3: 2004-2016 Concentration Growth Within Category



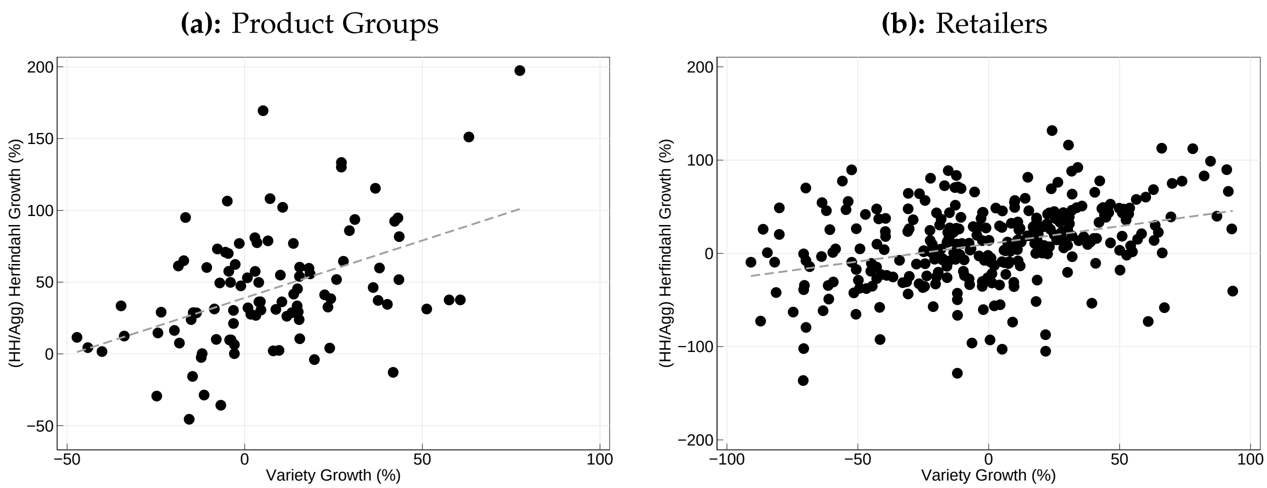
3.5 The Role of Product Churn

Interestingly, there is a common observable linking together the categories with the most rapid increases in niche ratios: they are also the categories that appear to have the fastest growth in the number of available products. We emphasize this relationship as it will be central to the mechanism in our model in Section 4 and its implications for welfare. As we noted above and elaborate in Appendix A.3, inference about the growth of aggregate product varieties in these data are highly sensitive to the treatment of products receiving trivial amounts of sales. In Figure 4a, however, we include products consumed by at least two households with total sales of at least $50 and plot the growth in each product group's niche ratio against the growth in that group's number of varieties. 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 ratios, with the relationship again significant at the 1 percent level.[17] The relationship between variety growth and the niche ratio becomes even steeper if we weight retailers by size.

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. For each household i that is observed in both t and $t+1$, we measure concentration of

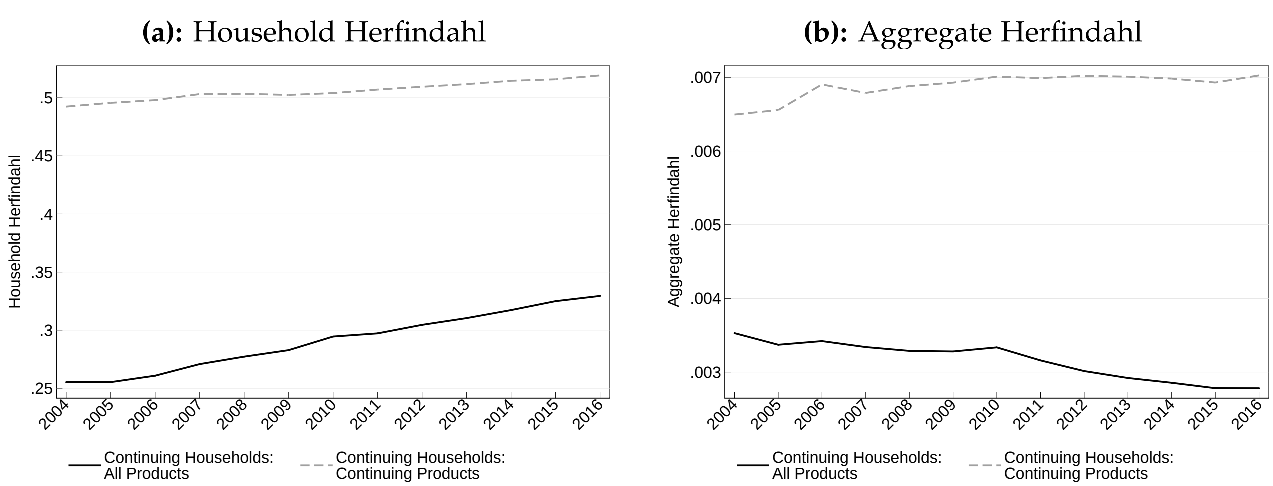
[17] To reduce the influence of outliers, we exclude retailers with absolute log variety changes above 2, which drops 6 out of 334 retailers. Results are similar for alternative thresholds. The panel of retailers is unbalanced, and growth rates for the remaining 328 retailers are calculated from their 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.

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



"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. 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. 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.

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



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.

4 Modeling the Rise of Niche Consumption

In this section, we develop a model that is able to match the rise of niche consumption documented in Section 3. We use the model to identify the key driving forces of this trend and to evaluate the resulting implications for welfare and markups.

The basic problem of an individual household follows Li (2019). Households have CES preferences which embed a love-of-variety but must pay a fixed cost per consumed product, which implies they only consume a subset of available products. We assume each household's taste for products, relative to their exogenous price, is distributed Pareto, which allows us to analytically characterize the concentration of each household's spending across products.

Our key theoretical contribution is the introduction of household heterogeneity into this framework. This innovation is crucial for matching the joint distribution of household and aggregate product spending and thus speaking to the rise of niche consumption. We do so by introducing a rank function in the basic household problem, that allows the preference ordering of products to differ across households while still allowing us to derive closed-form analytical expressions for aggregate spending patterns including the Aggregate Herfindahl. Using these analytical expressions, we confront the model with empirical trends from 2004-2016 and demonstrate 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.

Here, we offer simplified expressions for the key objects in the model. Interested readers can find detailed derivations of all analytical results in Appendix B.1.

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:

$$U_i = \left(\int_{k\in\Omega_i} \left(\gamma_{i,k} C_{i,k}\right)^{\frac{\sigma-1}{\sigma}} dk \right)^{\frac{\sigma}{\sigma-1}} - F \times \left(|\Omega_i|\right)^{\epsilon}, \tag{9}$$

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.[18]

[18]To ease notation, we do not include a household subscript i for each k, but importantly note that k is an arbitrary household-specific index of products, and so the same k may represent a different actual product for each different household. This is unimportant for the analysis of individual households, but will be crucial when we move to the aggregate analysis.

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:

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

where $y \geq b > 0$ and where we assume $\theta > 2\left(\sigma - 1\right)$. 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 $|\Omega_i|$. Household i will consume the set of goods with $\tilde{\gamma}_{i,k} \in [\tilde{\gamma}^*, \infty)$ for some $\tilde{\gamma}^* \geq b$.

The assumption, also in Li (2019), that price-adjusted tastes $\tilde{\gamma}_{i,k}$ follow a Pareto distribution, rather than tastes $\gamma_{i,k}$ themselves, warrants some additional discussion. When applying our static model at a given point in time, there is no substantive difference between assuming a distribution of prices, tastes, or price-adjusted tastes. However, when we apply our model to dynamic applications like the analysis of concentration trends, this assumption means that any changes in the distribution of prices must be accompanied by changes in the distribution of tastes in such a way that preserves the property that price-adjusted tastes remain Pareto distributed. We require this assumption for analytical tractability but nonetheless show in numerical results in Appendix B.2 that in our key counterfactual exercise in which prices adjust to changes in the number of varieties, the magnitude of the resulting required changes in tastes is small.

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

$$\begin{aligned} P_i = P &= \left(\int_{k\in\Omega_i} \left(\tilde{\gamma}_{i,k}\right)^{\sigma-1} dk\right)^{\frac{1}{1-\sigma}} \\ &= \underbrace{\left(1 + \frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}} b^{-1}}_{\text{Average Price}} \times \underbrace{\left(|\Omega_i|\right)^{\frac{1}{1-\sigma}}}_{\text{Variety Effects}} \times \underbrace{\left(\frac{|\Omega_i|}{N}\right)^{\frac{1}{\theta}}}_{\text{Selection Effects}}. \end{aligned} \tag{10}$$

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, 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-of-variety term in CES models, which decreases with the measure of consumed products and increases 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 increases in the share of available products that are consumed and decreases in the extent to which households prefer some products to others.

The price index (10) reduces to more standard expressions in special cases. For example, consider

$\theta \to \infty$, which implies that households value all products identically at b, i.e. $\tilde{\gamma}_{i,k} = b$ for all 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, so $\Omega_i = N$. In this case, the last term reduces to 1 as there are no selection effects and the average price term fully captures the impact of heterogeneity in the desirability of products.

The properties of the CES price index imply we can re-write equation (9) 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:

$$|\Omega_i| = |\Omega| = \left(\frac{\left(\frac{1}{\sigma-1} - \frac{1}{\theta}\right) \left(1 + \frac{1-\sigma}{\theta}\right)^{\frac{1}{1-\sigma}} N^{\frac{1}{\theta}}}{\tilde{F}\epsilon} \right)^{\left(\epsilon + \frac{1}{1-\sigma} + \frac{1}{\theta}\right)^{-1}}, \tag{11}$$

where $\tilde{F} = F/(bE)$ is a parameter that shifts with spending, aggregate prices, and variety costs.[19] Importantly, the optimal choice of varieties yields a "cutoff" taste $\tilde{\gamma}^*$ that satisfies:[20]

$$G\left(\tilde{\gamma}^*\right) = 1 - \frac{|\Omega|}{N}, \tag{12}$$

and the share of household $i's$ expenditure on variety k is then given by:

$$s_{i,k} = \begin{cases} \left(P\tilde{\gamma}_{i,k}\right)^{\sigma-1}, & \tilde{\gamma}_{i,k} > \tilde{\gamma}^* \\ 0, & \tilde{\gamma}_{i,k} \leq \tilde{\gamma}^*, \end{cases} \tag{13}$$

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

4.2 Household Herfindahls

Given equation (13), it follows that the Household Herfindahl $\mathcal{H}^{\text{HH}}$ will be equal for all i and can be written as:

$$\mathcal{H}_i^{\text{HH}} = \mathcal{H}^{\text{HH}} = \int_{k \in \Omega_i} \left(s_{i,k}\right)^2 dk = N \int_{\tilde{\gamma}_i^*}^{\infty} \left(P_i \tilde{\gamma}_{i,k}\right)^{2(\sigma-1)} dG\left(y\right) = \frac{(\eta+1)^2}{4\eta} \frac{1}{|\Omega|}, \tag{14}$$

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

[19] When $N = 1$, this expression is the same as that in Li (2019) after substituting in his special case for b.

[20] The fact that households only consume a subset of the varieties in our environment follows directly from our specification of the fixed cost. This same feature arises in Michelacci *et al.* (2019), though it is generated less mechanically in their setup. In their environment, consumers pay a cost to try a product and there is uncertainty as to whether they will like it.

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.

4.3 Testing the Model at the Household Level

How well does this model fit household spending data? If we condition on the number of products $|\Omega_i|$ purchased by household i, equation (13) yields a testable relationship between i's spending shares on particular products k and i's ranking of those particular products from $[0, N]$. For analytical tractability, our model assumes a continuum of products but in Appendix B.1.7 we provide formulas which convert the continuous relationship in equation (13) to a discretized version which we then use to test against the data. Specifically, we interpret the spending share on discrete product K as the total spending share by that household on all products k in the span of $[(K-1)/|\Omega_i|, K/|\Omega_i|]$ and obtain:

$$s_{i,K}^{|\Omega_i|} = (|\Omega_i|)^{-\frac{\eta+1}{2}} \left[K^{\frac{\eta+1}{2}} - (K-1)^{\frac{\eta+1}{2}} \right], \tag{15}$$

where we note that the spending share on the Kth discrete product is a function of both $|\Omega_i|$ and η. The share spent on the most preferred good (i.e. $K = 1$) simplifies to: $s_{i,1}^{|\Omega_i|} = (|\Omega_i|)^{-\frac{\eta+1}{2}}$.

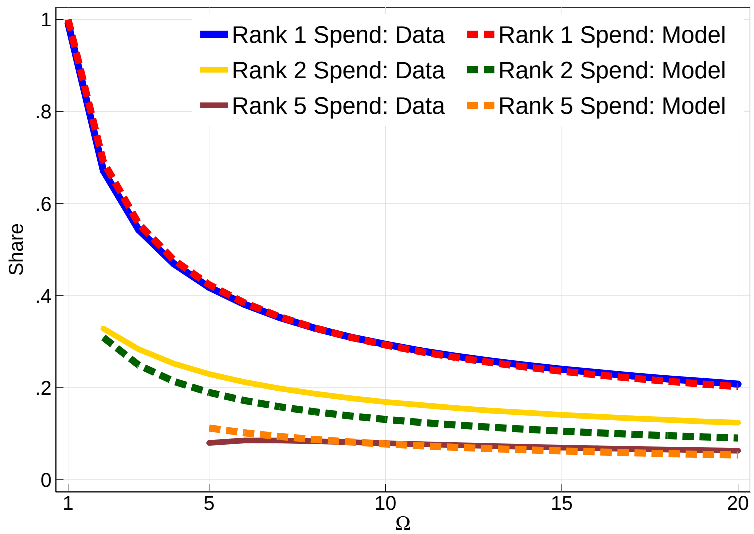
Panel (a) of Figure 6 plots the average spending shares in both the model and data on a household's top, second, and fifth ranked goods as a function of the total number of goods consumed by that household.[21] The solid lines measure the average of these shares in our data, taken across all households and weighted by their total spending, whereas the dashed lines simply plot the theoretical relationship implied by our model as shown in equation (15). For example, the solid and dashed lines in blue and red at the top of Figure 6 both equal 1 when $|\Omega| = 1$ at the far left of the plot, which must be true by construction. After all, if a household only buys one good, then that top good must by definition account for 100 percent of the spending. The far right side of the plot shows that for households who buy 20 products ($|\Omega| = 20$), the top product they purchase accounts for approximately 20 percent of their total spending in both the data and the model. The fact that the solid blue and dashed red lines nearly perfectly coincide for all values of $|\Omega|$ implies that our model of household spending does a great job capturing the degree to which households concentrate spending on their single most preferred product.

Next, the yellow solid and green dashed lines plot household spending shares on their second-ranked goods. This share is undefined when households only consume a single good, but ranges from about 0.3 to 0.1 as $|\Omega|$ ranges from 2 to 20. Households in the data spend a bit more on their second ranked good compared to what is predicted by the model, but the shares never deviate by more than a few percentage points. The solid brown and dashed orange lines show that the share spent on the

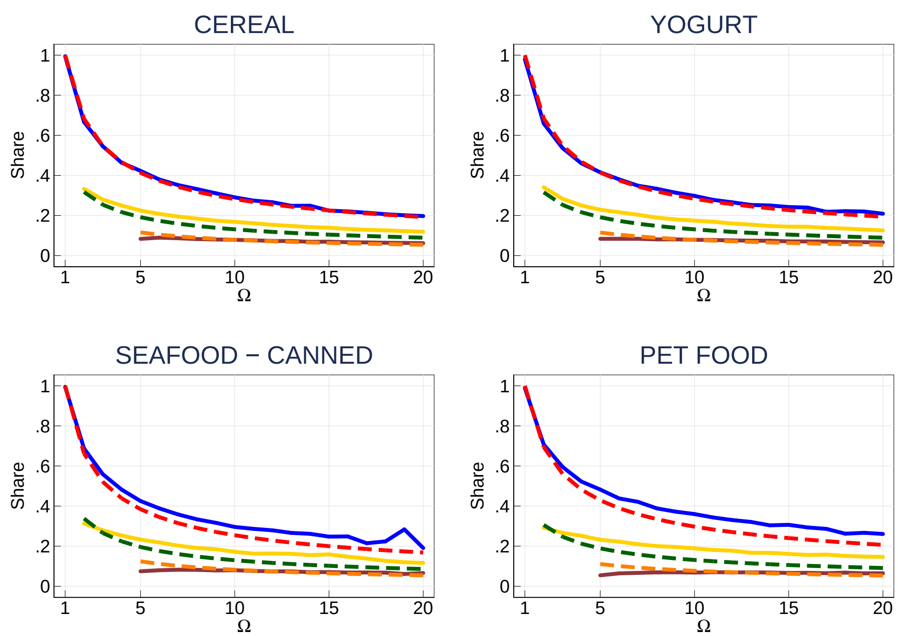
[21] We plot three ranks to preserve readability of the figure, but there is nothing special about these particular ranks. In Appendix Figure A6, we show that a focus on additional product ranks leads to similar conclusions.

Figure 6: Model Fit to Distribution of Spending by Product Rank for Households with Different Ω

(a): Average Spending on Different Ranked Items



(b): Category Spending on Different Ranked Items



fifth-ranked good, in both the data and the model, hovers quite close to 10 percent for all plotted values of $|\Omega|$.

Figure 6b repeats this exercise for the individual categories of cereal, yogurt, canned seafood, and pet food.[22] Although the model fits more closely for cereal and yogurt, the dashed lines are fairly

[22]We calibrate η for each category using equation (14) given the values of $\mathcal{H}^{\mathrm{HH}}$ and Ω. The lowest implied η in the plotted categories is 0.053 for pet food and the highest is 0.189 for seafood-canned. The η used for aggregate spending is 0.0672.

close to their solid counterparts in all cases. Thus, even though our model makes strong parametric assumptions in order to deliver analytical tractability, Figure 6 provides strong evidence that our model captures the key patterns of individual household spending we focus on in the data.[23]

4.4 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 by $j \in [0, N]$, and assume each household i assigns each product j a "rank" $r_{i,j}$, where lower ranks indicate higher price-adjusted tastes.[24] Households will consume all goods with $r_{i,j} \leq |\Omega|$.

We introduce the following rank function for each household i:

$$r_{i,j} = (1 - \alpha) j + \alpha x_{i,j}, \tag{16}$$

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 households and products. Without loss of generality, we assume $\alpha < 1/2$ and write:[25]

$$R(r) = \begin{cases} \frac{1}{2}\left(\frac{r}{N}\right)^2 \frac{1}{\alpha\,(1-\alpha)}, & 0 \leq r < \alpha N \\ \frac{r}{N}\frac{1}{1-\alpha} - \frac{1}{2}\frac{\alpha}{1-\alpha}, & \alpha N \leq r < (1-\alpha)\,N \\ -\frac{1}{2}\left(\frac{r}{N}\right)^2 \frac{1}{\alpha\,(1-\alpha)} + \frac{r}{N}\frac{1}{\alpha\,(1-\alpha)} - \frac{1}{2}\left(\frac{\alpha}{1-\alpha} + \frac{1-\alpha}{\alpha}\right), & (1-\alpha)\,N \leq r \leq N. \end{cases} \tag{17}$$

Note that this CDF satisfies the properties that $R(0) = 0$, $R(N) = 1$, $R(r)$ is continuous at $r = \alpha N$ and $r = (1-\alpha)\,N$, and $R(r)$ is monotonically increasing. Household i will consume good j if and only if

[23] Appendix A.5 derives additional testable implications of the household model and further corroborates its close fit to the data.

[24] Note that in contrast to the arbitrary household-specific product index k above, the product index j is common to all households.

[25] Replacing α with $1 - \alpha$ in all instances in equation (17) 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.

$R(r_{i,j}) \leq |\Omega|/N$.

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.[26] 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 the analytical expressions that follow, we thus impose this parameter restriction for the remainder of the analysis.

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

$$s_{i,j} = P^{\sigma-1}\tilde{\gamma}_{i,j}^{\sigma-1} = (Pb)^{\sigma-1}\left(R\left(r_{i,j}\right)\right)^{-\frac{\sigma-1}{\theta}} = \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\left(R\left(r_{i,j}\right)\right)^{\frac{\eta-1}{2}}, \tag{18}$$

if $R\left(r_{i,j}\right) \leq |\Omega|/N$, and zero otherwise. To determine the products for which the share $s_{i,j}$ in equation (18) jumps 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 (17) under the assumption that $0 \leq \frac{|\Omega|}{N} < \frac{\alpha}{2(1-\alpha)}$, we get:

$$r^* = \left(2\alpha\left(1-\alpha\right)|\Omega|N\right)^{\frac{1}{2}}. \tag{19}$$

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 not consumed by any household. 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 with 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 $x_{i,j} = 0$) yields rank r^* for the household with that zero draw.

[26] More 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.

Solving for this cutoff, $j^* = r^* / (1-\alpha)$, we get:

$$j^* = \left(\frac{2\alpha|\Omega|N}{1-\alpha}\right)^{\frac{1}{2}}. \tag{20}$$

Importantly, since $r_{i,j}$ is strictly increasing in j, all goods with $j \leq j^*$ will have positive aggregate sales and all goods with $j > j^*$ will have zero aggregate sales. Finally, substituting in the definition of the rank function from equation (16) into the expression (19), and using the definition of j^* in equation (20), we can write the highest value or cutoff random draw x_j^* that yields positive consumption of j as:

$$x_j^* = \frac{1-\alpha}{\alpha}\left(j^* - j\right). \tag{21}$$

4.5 The Aggregate Herfindahl

We now use equations (18) and (21) to integrate spending shares across households i to get the aggregate spending share on good j:

$$\begin{aligned} s_j &= \frac{1}{\int_i Edi}\int_i Es_{i,j}di = \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\int_0^{x_j^*}\left(R\left((1-\alpha)j+\alpha x\right)\right)^{\frac{\eta-1}{2}}\frac{dx}{N} \\ &= \frac{\eta+1}{\eta j^*}\left(1-\left(\frac{j}{j^*}\right)^{\eta}\right). \end{aligned} \tag{22}$$

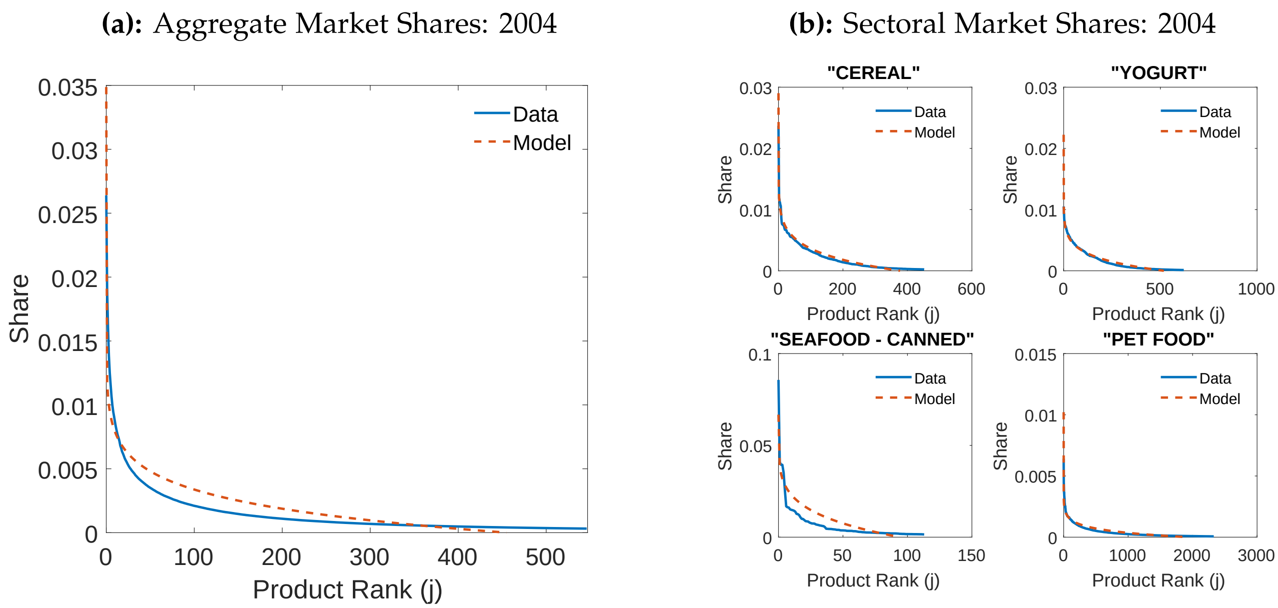
Using equations (20) and (22), we immediately obtain the Aggregate Herfindahl:

$$\begin{aligned} \mathcal{H}^{\text{Agg}} &= \int_{j=0}^{j^*} s_j^2 dj = \left(\frac{\eta+1}{\eta j^*}\right)^2 \int_{j=0}^{j^*}\left(1-\left(\frac{j}{j^*}\right)^{\eta}\right)^2 dj \\ &= \frac{2(\eta+1)}{2\eta+1}\left(\frac{1}{2\tilde{N}|\Omega|}\right)^{\frac{1}{2}}, \end{aligned} \tag{23}$$

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 $\mathcal{H}^{\text{HH}}$ and $\mathcal{H}^{\text{Agg}}$ and implies an important role for increases in $\tilde{N}$.

How well do these model-based relationships fit aggregate sales distributions in the data? To assess this, we start by measuring $|\Omega_c|$ in the data as the average number of products consumed per household within a category c using the same weights as were used in equations (3)-(5). We then solve for the two remaining free parameters, η_c and $\tilde{N}_c$, to match $\mathcal{H}_c^{\text{HH}}$ and $\mathcal{H}_c^{\text{Agg}}$ in equations (14) and (23). Figure 7 then plots the market share distribution across products implied by our model in equation (22) (the red dashed line) against the actual market share distribution in the data (the solid blue line)

Figure 7: Model Fit on Aggregate Category Data



in 2004.[27] We weight across categories and 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 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.[28]

4.6 Elasticities of Demand, Markups, and Aggregate Profits

The previous sections develop analytical expressions for our key empirical objects: $\mathcal{H}^{\mathrm{HH}}$, $\mathcal{H}^{\mathrm{Agg}}$, and $|\Omega|$. We will show that these expressions can be used to draw important conclusions about the forces driving the rise of niche consumption, but we defer this analysis until Section 4.7. First, in this subsection, we additionally develop an expression for the ratio of aggregate revenues to costs, what we refer to as the "aggregate markup". The aggregate markup is useful on its own as a gauge of market power, but further, we will use this expression to calibrate our model and quantify the welfare implications of the rise in niche consumption.

In typical CES environments, the elasticity of demand and markups are fully determined by the exogenous elasticity of substitution σ. By contrast, the elasticity of demand in our model depends both

[27] We concentrate on a single year since our model is static. Results are similar if we instead use data from 2016 or pool all years.

[28] Appendix Section B.3 offers a version of our model that builds from a linear demand system as in Melitz and Ottaviano (2008), rather than the CES demand system assumed above. In this case, we also obtain analytical expressions for key statistics such as Household and Aggregate Herfindahls. Most of the qualitative inferences from confronting the model with the data are the same across both model specifications. The CES model is our benchmark in large part because the match between its predictions and the data, as explored in Figure 7, strike us as more compelling.

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 heterogeneity in demand elasticities across products and in their resulting markups. This also implies that existing estimates of the elasticity of substitution, such as those offered in 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:

$$s_j = \frac{1}{N} \int_0^{x_j^*} s_{i_x,j} dx, \tag{24}$$

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 (24) with respect to p_j to get:

$$\frac{\partial s_j}{\partial p_j} = \frac{1}{N} \left(\int_0^{x_j^*} \frac{\partial s_{i_x,j}}{\partial p_j} dx + s_{i_{x_j^*},j} \frac{\partial x_j^*}{\partial p_j} \right), \tag{25}$$

where the right hand side of equation (25) follows from Leibniz's rule. The first term can be solved using equation (18) as:

$$\frac{\partial s_{i_x,j}}{\partial p_j} = (1 - \sigma) \frac{s_{i_x,j}}{p_j}, \tag{26}$$

where we take the aggregate price index P as fixed. Moving on to the second term, we can evaluate equation (18) at the marginal household with taste x_j^* to get:

$$s_{i_{x_j^*},j} \frac{\partial x_j^*}{\partial p_j} = \frac{\eta + 1}{2} N^{\frac{\eta-1}{2}} |\Omega|^{-\frac{\eta+1}{2}} \left(R\left(r^*\right) \right)^{\frac{\eta-1}{2}} \frac{\partial x_j^*}{\partial p_j}. \tag{27}$$

Substituting equations (26) and (27) back into equation (25), we get:

$$\frac{\partial s_j}{\partial p_j} = (1 - \sigma) \frac{1}{p_j} \frac{1}{N} \int_0^{x_j^*} s_{i_x,j} dx + \frac{1}{N} s_{i_{x_j^*},j} \frac{\partial x_j^*}{\partial p_j} = (1 - \sigma) \frac{s_j}{p_j} + \frac{\eta + 1}{2N|\Omega|} \frac{\partial x_j^*}{\partial p_j}. \tag{28}$$

Our analytical expressions thus far rely on the assumption that the full distribution of price-adjusted tastes in our model is given exogenously as Pareto. However, since we have a continuum of products, we assume that the influence of an infinitesimal price change on the overall distribution of demand is marginal and so our analytical expressions continue to hold.[29] To compute $\partial x_j^* / \partial p_j$, we

[29] In Appendix B.2 we use numerical simulations which do not require any distributional assumption on tastes to verify

start with the relationship:

$$R\left(\left(1-\alpha\right)j+\alpha x_{i,j}\right)=1-G\left(\frac{\gamma_{i,j}}{p_j}\right)=b^{\theta}\gamma_{i,j}^{-\theta}p_j^{\theta}, \tag{29}$$

and totally differential the left- and right-hand sides for each product j. Evaluating the resulting expression at $x_{i,j}=x_j^*$, $r_{i,j}=r^*$, and $\tilde{\gamma}_{i,j}=\tilde{\gamma}^*$ yields:

$$\frac{dx_j^*}{dp_j}=-\frac{\theta}{j^*}|\Omega|N\frac{1}{p_j}. \tag{30}$$

Inserting this into equation (28), we have:

$$\frac{\partial s_j}{\partial p_j}=\left(1-\sigma\right)\frac{s_j}{p_j}-\frac{\eta\theta}{2\left(1-\left(\frac{j}{j^*}\right)^{\eta}\right)}\frac{s_j}{p_j}. \tag{31}$$

Equation (31) implies that product j's price elasticity of demand ε_j can be written as:

$$\varepsilon_j=1-\frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j}=\underbrace{\sigma}_{\text{Intensive Margin}}+\underbrace{\left(1-\left(\frac{j}{j^*}\right)^{\eta}\right)^{-1}\left[\theta/2-\left(\sigma-1\right)\right]}_{\text{Extensive Margin}}\quad>\quad\sigma. \tag{32}$$

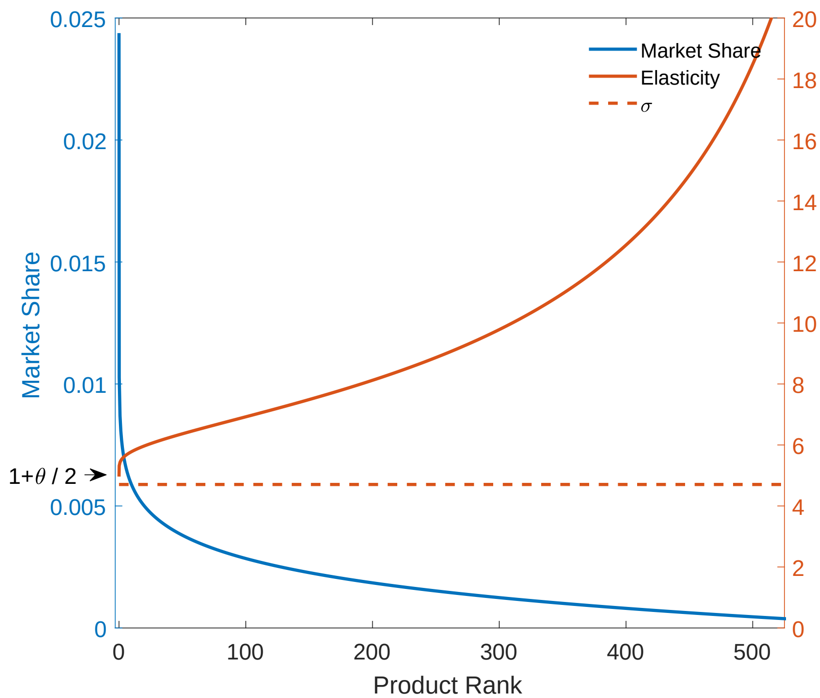
In addition to the standard intensive margin term σ, there is a strictly positive contribution from the extensive margin, since lowering the price of a product can induce new households to start consuming the product. Low j or "mass-market" products are consumed by many households, so the intensive margin is relatively more important for them. High j or "niche" products are consumed by few households, so the extensive margin is relatively more important for those goods. As a result, the elasticity of demand increases as market share falls.[30]

Figure 8 plots aggregate product market shares s_j, the elasticity of demand ε_j, and the elasticity of substitution σ as a function of the product rank j, using a version of the model economy that we calibrate as described below. Elasticities of demand rise as the product rank grows from the low values associated with large market shares to the high values associated with niche products. We note that despite the CES structure, the elasticity of demand generically differs from the elasticity of substitution σ. As $j\to j^*$ and a product approaches the point where it is dropped from the aggregate consumption bundle, the elasticity approaches infinity, i.e. $\varepsilon\to\infty$. The markup μ_j then be written (in gross terms)

that the approximation that follows is highly accurate.

[30]Interestingly, 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.

Figure 8: Elasticity of Demand for Good j



as:

$$\mu_j \quad = \quad \frac{\varepsilon_j}{\varepsilon_j - 1} = \frac{\sigma + \frac{\theta(\eta+1)}{2j^* s_j}}{\sigma + \frac{\theta(\eta+1)}{2j^* s_j} - 1}, \tag{33}$$

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

The aggregate markup μ^{Agg} is equal to the ratio of aggregate sales to aggregate costs. Using equations (22) and (32), it can be written as:

$$\mu^{\text{Agg}} = \frac{\int_0^{j^*} s_j dj}{\int_0^{j^*} s_j \frac{\varepsilon_j - 1}{\varepsilon_j} dj} = \left[\frac{\theta + (\sigma-1)^2}{\sigma^2} - \frac{1}{2} \frac{\eta \theta^2}{\sigma^2} \left(\frac{\eta+1}{2+\theta} \right) \times {}_2F_1 \left(1, \frac{1}{\eta}; 1 + \frac{1}{\eta}; \frac{2\sigma}{2+\theta} \right) \right]^{-1}, \tag{34}$$

where ${}_2F_1(\cdot)$ is the hypergeometric function.[31] 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.[32]

---

[31] The 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 $n = 0$.

[32] We 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 B.2, 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.

4.7 Understanding the Empirical Trends

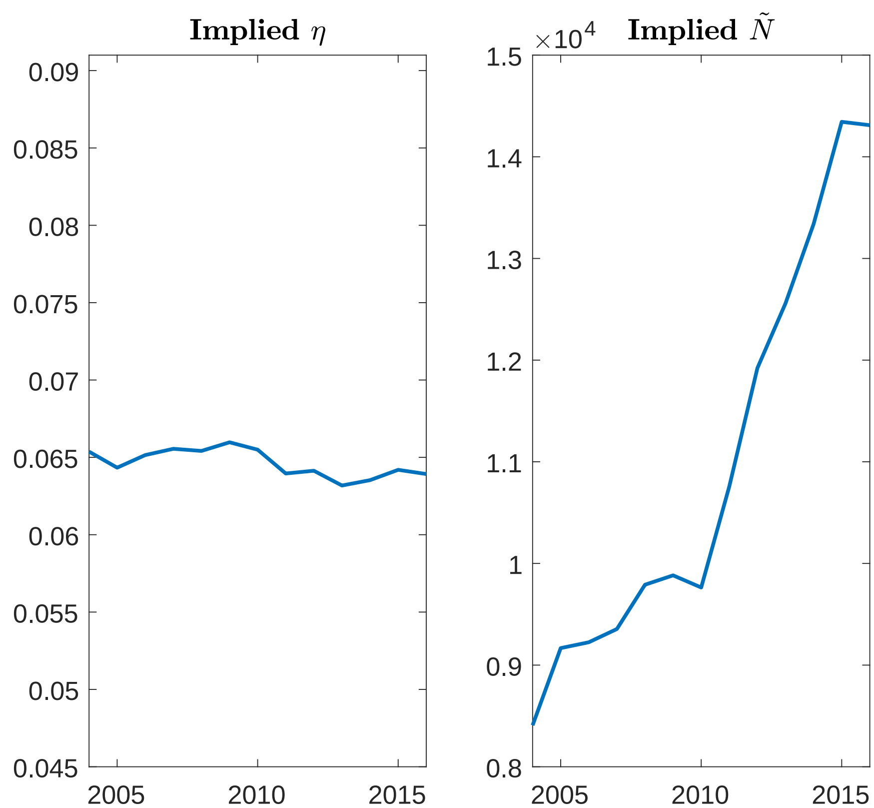
We now confront our model with concentration measures and other moments from the data to infer which structural forces led to the rise in niche consumption. Collecting previous results, our model implies that:

$$\mathcal{H}^{\text{HH}} = \frac{(\eta+1)^2}{4\eta}\frac{1}{|\Omega|} \tag{35}$$

$$\mathcal{H}^{\text{Agg}} = \frac{2(\eta+1)}{(2\eta+1)}\left(\frac{1}{2\tilde{N}|\Omega|}\right)^{\frac{1}{2}}. \tag{36}$$

Since $\mathcal{H}^{\text{HH}}$, $\mathcal{H}^{\text{Agg}}$, and $|\Omega|$ are directly observable in the data, this produces a system of two equations – (35) and (36) – that can be solved to determine η and $\tilde{N}$ for each year. Figure 9 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.

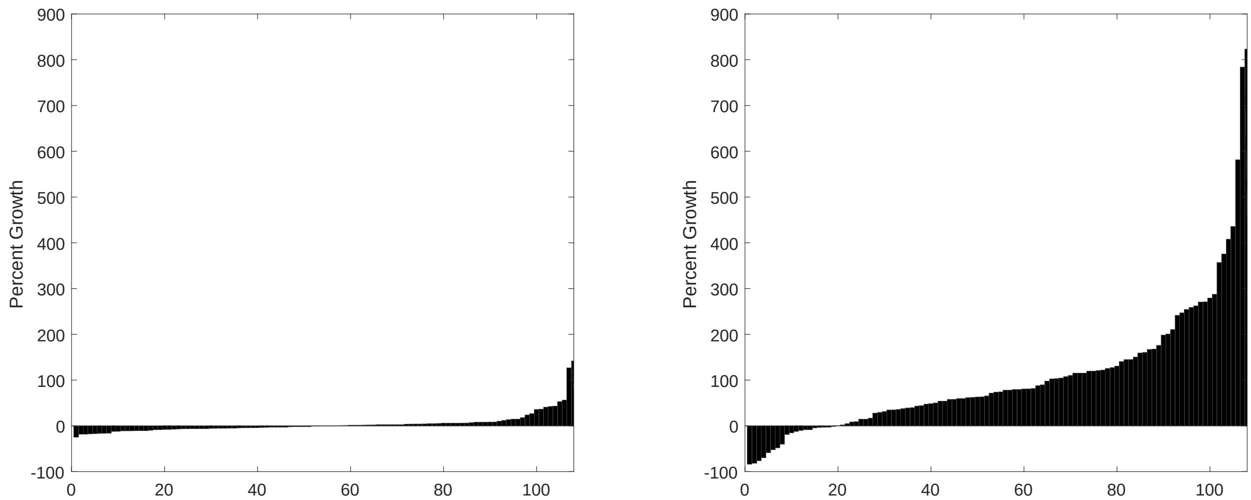
Figure 9: Implied Drivers of the Aggregate Rise of Niche Consumption



As above, we can interpret our model as applying at each individual sector and apply equations (35) and (36) to data on concentration trends and varieties consumed sector-by-sector. Rather than generating a single time series for η and a single time series for $\tilde{N}$ for the aggregate economy, as plotted in Figure 9, this allows us to generate time series for these parameters for each sector. Figure 10a plots the distribution of implied growth (in percent terms) from 2004-2016 of η for each sector, and Figure 10b plots the same for the implied growth of $\tilde{N}$.

Figure 10: Implied Drivers of the Sector-Level Rise of Niche Consumption

(a): Sectoral Distribution of Growth in η **(b):** Sectoral Distribution of Growth in $\tilde{N}$



It turns out that our aggregate results represent 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.

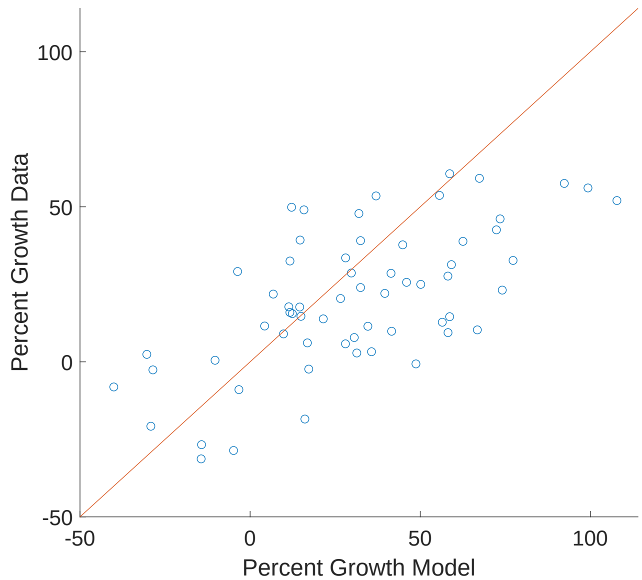
Increases in $\tilde{N}$ can arise from increases in the importance of idiosyncratic taste shocks α or in the number of available varieties N. Changes in α are straightforward to interpret, since α is simply an exogenous parameter governing preference heterogeneity. Our empirical results, however, show 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.[33]

How do these implied growth rates for the number of available varieties in each sector compare to

[33] 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.

estimates from the data? As we have emphasized, measuring the total number of varieties purchased in the data can only be done with substantial noise (see Appendix A.3). However, Figure 11 plots the growth in sectoral varieties with at least \$100 in annual spending in the data against the model-implied growth in varieties for each sector j^*, the values shown in Figure 10b. The values are broadly consistent and clustered around the 45 degree line, which provides some direct confirmation of the model's inference.

Figure 11: Consumed Variety (j^*) Growth: Model vs. Data



Finally, we note that while increases in N are necessary to fit the empirical patterns we emphasize, they are not sufficient. The implied increase in N matches the divergence between Household and Aggregate Herfindahls given the empirical decline in $|\Omega|$, but as shown in equation (11), increases in N on their own would counterfactually lead to increases in $|\Omega|$. This means that additional forces are required in order to fully fit the empirical trends.

Equation (11) shows that if N increases, declines in $|\Omega|$ must reflect declines in measured real expenditures $\left(E\left(1+(1-\sigma)/\theta\right)^{\frac{1}{1-\sigma}} b\right)$, increases in effective costs per number of products consumed (F or ϵ), or declines in an "effective curvature" of utility term $\left(\frac{1}{1-\sigma}-\frac{1}{\theta}\right)$. 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.[34] 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. Rather, our

[34] While 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.

quantitative results are largely driven by the change in N.

4.8 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 look at combinations of changes in N, σ, θ, and the fixed costs of variety that fit the rise in niche consumption. One difficulty is that while we can identify $\eta = 1 - 2(\sigma - 1)/\theta$, we cannot separately identify σ and θ and, as shown above, our model implies we cannot simply use existing estimates of the elasticity of demand as a proxy for σ. To get around this problem, we assume the aggregate markup equals 1.15 in 2004, or μ^{Agg}. We then use equation (34), together with our implied value for η, to calculate the implied values $\sigma = 4.7$ and $\theta = 7.9$.[35] Below, we consider robustness to this chosen target for μ^{Agg}.

Having calibrated the key parameter values, we use the model to explore the welfare implications of several different counterfactuals. We begin by evaluating the implications of an increase in N equal to what we plotted in Figure 9, 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. Row (i) of 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 7.0 percent, or 0.56 percent per year. That is:

$$U_{2016} = \frac{E}{P_{N_{2016}}} - F \times \left(|\Omega_{N_{2016}}|\right)^{\epsilon} = 1.070 \times \frac{E}{P_{N_{2004}}} - F \times \left(|\Omega_{N_{2004}}|\right)^{\epsilon}, \tag{37}$$

where we change N and calculate the endogenous change in P and $|\Omega|$, but hold fixed all other parameters. Next, we perform similar counterfactuals but include changes in ϵ or F to also match the change in $|\Omega|$. As shown in rows (ii) and (iii) of Table 1, implied welfare growth remains large at 0.46-0.47 percent, so that the effect of increasing fixed costs is quantitatively small relative to the increase in N. Finally, we add changes in either σ or in θ to match the small implied decline in η, plotted in Figure 9. As shown in rows (iv) and (v), these additional changes have almost no effect on the results.

Table 1 also decomposes the welfare gains into four sources (that needn't sum exactly due to non-linearities). First, there are "Gains from Selection", which come from the third term in the ideal price index in equation (10) and emerge when $|\Omega|/N$ decreases, implying that households consume products better suited to their particular tastes. These gains are the most important quantitatively. For

[35] This value for the aggregate markup is the preferred value in Edmond *et al.* (2018) and is close to other papers that employ a variety of methodologies. Other parameters are not important for our quantitative conclusions, but we set $\alpha = 0.36$, $\epsilon = 2$, $E = 35$, $b = 1$, and $F = 0.055$. E is set to match average household category expenditures. Given b and ϵ, we choose F to match $|\Omega|$.

example, in row (i), 0.52 of the 0.56 percent total annual welfare gains come from the "Gains from Selection" term. Second, changes in $|\Omega|$ show up as standard love-of-variety effects on welfare, even if selection effects $|\Omega|/N$ are held constant. Whether these "Gains from Variety" are positive, as in row (i) which does not match the observed decline in the number of varieties consumed by each household $|\Omega|$, or negative, as in the others which do match this decline in $|\Omega|$, gains or losses from variety are much smaller in magnitude than those from selection. Third, there are relatively minor welfare implications of changes in fixed costs brought about by changes in $|\Omega|$, F, or ϵ. Finally, in specifications where we change σ or θ, there is a trivially small impact from changes in the "Average Price", which captures the changing price of purchasing a bundle of all available varieties. Overall, the conclusions from Table 1 are simple: the rise of niche consumption is associated with substantial welfare gains, and these arise almost entirely from greater selection as N increases.

If data were available at the household level to calculate a spending-weighted variety correction as in Feenstra (1994), one could properly recover welfare growth, as we show in Appendix Section B.1.8. However, the data necessary to confidently construct such measures at the individual household-level is seldom available. For instance, variety churn for the typical household in our data is incredibly large and variable and so such an exercise would produce a huge range of potential conclusions. Our approach instead is to use a model that points us to easily observable and relatively stable values from the data and specifies how to combine those values to assess changes in welfare. The approach has the additional merit of decomposing welfare gains into the terms emphasized in Table 1 and allowing us to run counterfactuals.

If, instead, one simply viewed our data through the lens of a representative household model with CES preferences, one could calculate a variety correction on aggregate spending following Feenstra (1994). But, as is shown in Appendix Section B.1.8, the calculated welfare gains in this case would not coincide with the true household welfare gains. The Feenstra correction is derived in the context of a CES framework. Our model assumes that households individually have CES preferences, but the heterogeneity across households means that the aggregate economy in our model does not admit a representative household representation with CES preferences.

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.

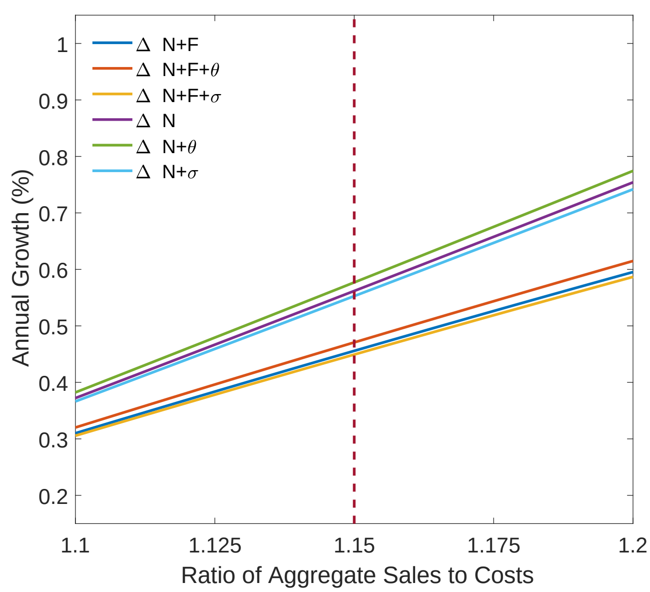
The welfare effects of increased product selection do depend importantly on θ, which as noted above, we pin down by targeting an aggregate markup of 1.15 together with the initial η value in Figure 9. Figure 12 shows how implied welfare growth changes under alternative calibrations for the initial aggregate markup. The vertical dashed red line corresponds to our baseline calibration and the "Total" numbers in the first column of Table 1. For example, the yellow line intersects the red dashed

Table 1: Annualized Welfare Growth (Compensating Expenditures)

		Total	Average Price	Gains from Variety	Gains from Selection	Fixed Costs
		$d \ln E$	$\left(\frac{1+\theta-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}}$	$(\lvert\Omega\rvert)^{\frac{1}{1-\sigma}}$	$(\lvert\Omega\rvert/N)^{\frac{1}{\theta}}$	$-F \times (\lvert\Omega\rvert)^{\epsilon}$
(i)	$\uparrow N$	0.56%	0%	0.08%	0.52%	-0.05%
(ii)	$\uparrow N, \uparrow \epsilon$	0.47%	0%	-0.13%	0.62%	-0.02%
(iii)	$\uparrow N, \uparrow F$	0.46%	0%	-0.13%	0.62%	-0.04%
(iv)	$\uparrow N, \uparrow F, \uparrow \sigma$	0.45%	-0.001%	-0.14%	0.62%	-0.03%
(v)	$\uparrow N, \uparrow F, \downarrow \theta$	0.47%	-0.003%	-0.13%	0.63%	-0.04%

line at 0.45 and corresponds to the counterfactual in row (iv) which fits the data by changing N, F, and σ. The purple line intersects at 0.56 and corresponds to the experiment where we only increase N, as in row (1) of the table. Figure 12 shows that welfare effects remain large for a wide range of markup calibration choices.

Figure 12: Robustness of Welfare Calculations to Calibration of θ and σ



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 2.[36] Cate-

[36]We maintain $E = 35$ in all categories but this is a normalization without loss of generality since we recalibrate F for each category to target that category's $|\Omega|$. More substantively, we assume each sector's aggregate markup also equals 1.15. If we instead impose a common σ and only use heterogeneity in θ_c to match sectoral variation in η_c, we find similar welfare

gories such as coffee, snacks, and soup all rank in the top 10 for welfare 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 leads to welfare losses, in eggs, cottage cheese, and frozen vegetables.

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	1.40	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	1.30	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	0.30
(10)	Pizza	1.15	0.78	0.94	(42)	Cosmetics	0.41	0.92	0.50
(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	0.80	(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	1.00	-0.62	0.50	(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	1.40	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	0.60	(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	-0.30
(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	0.30	0.29	(64)	Photographic Supplies	-1.83	-6.48	-3.07

4.9 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 very small. 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^* in equation (33), but changing them leaves the aggregate markup exactly constant, since j^* drops out of equation (34). More intuitively, this result arises from two opposing forces which exactly cancel when j^* changes. On the one hand, the jth good in an economy with a low j^* is closer to being the marginal consumed good and will therefore have a lower markup then the jth good in an economy with a high j^*. This can be seen in equation (32), 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.

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.

On the other hand, in an economy with greater product choice, the high-markup products account for a smaller share of aggregate spending. This competitive force from increasing N reduces aggregate markups. Equation (34) shows that in the aggregate, these opposing forces exactly cancel and the ratio of total revenues to total costs, or the aggregate markup, remains unchanged. In specifications where we also change σ or θ, aggregate markups are no longer exactly fixed but resulting changes are tiny, rising by 0.02 percentage points if we vary θ to hit the change in η and falling by 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. More generally, echoing the arguments in Syverson (2018) and Berry *et al.* (2019), our environment demonstrates that depending on what forces drive changes in concentration, it is possible for aggregate market power to concurrently increase or to decrease.

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 changes in θ (Column 1) or by changes in σ (Column 2). Overall, the conclusions mirror that from the aggregate analysis. 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 modest declines in markups in photographic supplies, records and tapes, and lightbulbs. There are small increases in markups for coffee, soaps, tea, and cosmetics.

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 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

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	0.10	(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	0.00	(47)	Pet Care	0.09	-0.01
(16)	Breakfast Food	-0.03	0.00	(48)	Crackers	0.10	-0.01
(17)	Pet Food	-0.03	0.00	(49)	Candy	0.12	-0.02
(18)	Bread and Baked Goods	-0.03	0.00	(50)	Laundry Supplies	0.12	-0.01
(19)	Canned and Bottled Juice Drinks	-0.02	0.00	(51)	Frozen Vegetables	0.12	-0.02
(20)	Cereal	-0.02	0.00	(52)	Baking Supplies	0.12	-0.02
(21)	Liquor	-0.02	0.00	(53)	Spices, Seasonings, and Extracts	0.13	-0.02
(22)	Prepared Frozen Foods	-0.01	0.00	(54)	Jams and Jellies	0.13	-0.02
(23)	Packaged Deli Meats	-0.01	0.00	(55)	Medications	0.14	-0.02
(24)	Hair Care	0.00	0.00	(56)	Oral Hygiene	0.15	-0.02
(25)	Frozen Meats and Seafood	0.00	0.00	(57)	Packaged Milk	0.16	-0.02
(26)	Wrapping Materials And Bags	0.01	0.00	(58)	Yogurt	0.18	-0.02
(27)	Carbonated Beverages	0.02	0.00	(59)	Beer	0.20	-0.03
(28)	Soup	0.02	0.00	(60)	Household Cleaners	0.22	-0.03
(29)	Skin Care Preparations	0.03	0.00	(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

trends, and led to welfare gains from households being able to consume a subset of products that better satisfied their tastes. This welfare effect is not found in standard statistics such as the price indices produced by national statistical agencies. 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 and the measurement of consumer welfare.

References

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

Alwin, D. and Tufis, P. (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, E., Rebelo, S. and Wong, A. (2018). *Markups Across Space and Time*. Working Paper 24434, National Bureau of Economic Research.

Anderson, S., de Palma, A. and Thisse, J. (1987). The ces is a discrete choice model? *Economics Letters*, **24** (2), 139–140.

Argente, D., Lee, M. and Moreira, S. (2018a). How do firms grow? the life cycle of products matters.

—, — and — (2018b). Innovation and product reallocation in the great recession. *Journal of Monetary Economics*, **93**, 1–20.

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, **98** (5), 1998–2031.

Atkin, D., Faber, B. and Gonzalez-Navarro, M. (2018). Retail globalization and household welfare: Evidence from mexico. *Journal of Political Economy*, **126** (1), 1–73.

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

Autor, D. H., Katz, L. F. and Kearney, M. S. (2006). The polarization of the u.s. labor market. *The American Economic Review*, **96** (2).

Berger, D. and Vavra, J. (2015). Consumption dynamics during recessions. *Econometrica*, **83** (1).

— and — (2019). Shocks vs. responsiveness: What drives time-varying dispersion? *Journal of Political Economy*.

Berry, S. T., Gaynor, M. and Morton, F. S. (2019). *Do Increasing Markups Matter? Lessons from Empirical Industrial Organization*. Working Paper 26007, National Bureau of Economic Research.

Bertrand, M. and Kamenica, E. (2018). *Coming apart? Lives of the Rich and the Poor Over Time*. Tech. rep.

Broda, C. and Weinstein, D. W. (2004). Variety growth and world welfare. *American Economic Review*, **94** (2), 139–144.

Brynjolfsson, E., Hu, Y. and Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, **49** (11), 1580–1596.

Burstein, A., Grassi, B., Carvalho, V. *et al.* (2019). Bottom-up markup fluctuations. In *2019 Meeting Papers*, 505, Society for Economic Dynamics.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, **98** (4), 1707–21.

Davis, D. R., Dingel, J. I., Monras, J. and Morales, E. (2017). How segregated is urban consumption?

De Loecker, J. and Eeckhout, J. (2017). *The Rise of Market Power and the Macroeconomic Implications*. Working Paper 23687, National Bureau of Economic Research.

Dube, J.-P., Hitsch, G. and Rossi, P. (2018). Income and wealth effects on private-label demand: Evidence from the great recession. *Marketing Science*.

Edmond, C., Midrigan, V. and Xu, D. Y. (2018). *How costly are markups?* Tech. rep., 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.

— and Weinstein, D. E. (2017). Globalization, markups, and us welfare. *Journal of Political Economy*, **125** (4), 1040–1074.

Furman, J. and Orszag, P. (2015). A firm-level perspective on the role of rents in the rise in inequality. *Presentation at "A Just Society" Centennial Event in Honor of Joseph Stiglitz Columbia University*, **16**.

Gentzkow, M., Shapiro, J. M. and Taddy, M. (2017). *Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech*. Working Paper 22423, National Bureau of Economic Research.

Gopinath, G. and Itskhoki, O. (2010). Frequency of price adjustment and pass-through. *The Quarterly Journal of Economics*, **125** (2), 675–727.

Handbury, J. and Weinstein, D. E. (2014). Goods prices and availability in cities. *The Review of Economic Studies*, **82** (1), 258–296.

Hausman, J. A. (1996). *Valuation of New Goods under Perfect and Imperfect Competition*, University of Chicago Press, pp. 207–248.

Hottman, C. J., Redding, S. J. and Weinstein, D. E. (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.

Levin, A. and Yun, T. (2008). Reconsidering the microeconomic foundations of price-setting behavior. *Unpublished Manuscript, Board of Governors of the Federal Re serve System*.

Li, N. (2019). An engel curve for variety. *Review of Economics and Statistics*, pp. 1–45.

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *The review of economic studies*, **75** (1), 295–316.

Michelacci, C., Paciello, L. and Pozzi, A. (2019). The extensive margin of aggregate consumption demand. *Mimeo*.

Neiman, B. and Vavra, J. (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*. Tech. rep.

Piketty, T., Saez, E. and Zucman, G. (2016). *Distributional National Accounts: Methods and Estimates for the United States*. Working Paper 22945, National Bureau of Economic Research.

Quan, T. W. and Williams, K. R. (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 Weinstein, D. E. (2016). *Measuring aggregate price indexes with demand shocks: Theory and evidence for CES Preferences*. Tech. rep., National Bureau of Economic Research.

Rossi-Hansberg, E., Sarte, P.-D. and Trachter, N. (2018). *Diverging trends in national and local concentration*. Tech. rep., National Bureau of Economic Research.

Stroebel, J. and Vavra, J. (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 to:
The Rise in Niche Consumption

Brent Neiman
University of Chicago

Joe Vavra
University of Chicago

October 2020

Section A of this appendix is focused on our data and empirical results while Section B elaborates on our model and theoretical results.

Appendix A. Data Appendix

We start this section of the appendix with Subsection A.1, which offers a detailed description of the Nielsen Homescan dataset, and Subsection A.2 compares the spending growth in these data to that in other datasets. Subsection A.3 then discusses the difficulty of measuring the number of aggregate varieties in these data and demonstrates the sensitivity of such measures to the treatment of products with small spending shares, while Subsection A.4 corroborates that our results on aggregate concentration are not inconsistent with concentration measures calculated using census data. Finally, we conclude with Subsection A.6, which collects a number of additional empirical results.

A.1 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

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

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

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. 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 *et al.* (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

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

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 *et al.* (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. 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 concentration trends that we document.

Figure A1: Online Spending Shares

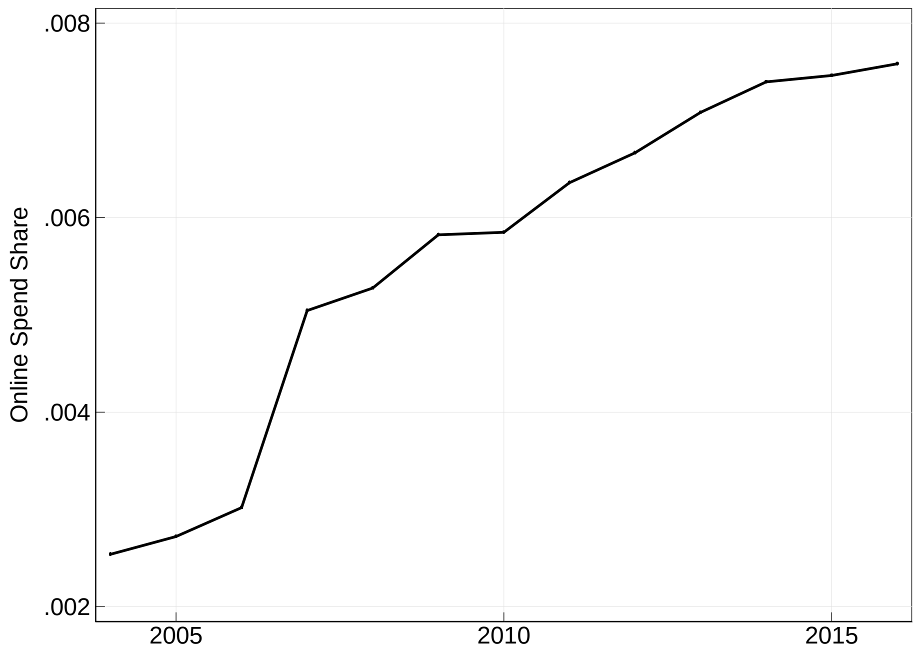
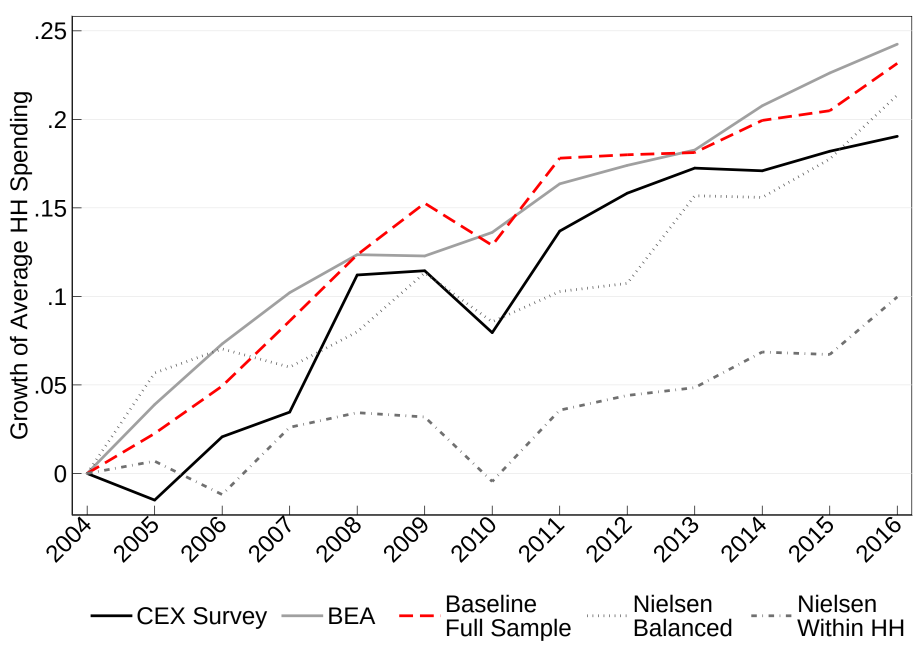


Figure A2: Household Spending in Nielsen vs. Consumer Expenditure Survey



A.2 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.[5]

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 U.S., 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 A13 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

[4] Online 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 does not classify as online spending the shopping with traditional retailers that happens to occur through their websites, such as spending at Walmart.com.

[5] It 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 consumer 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.

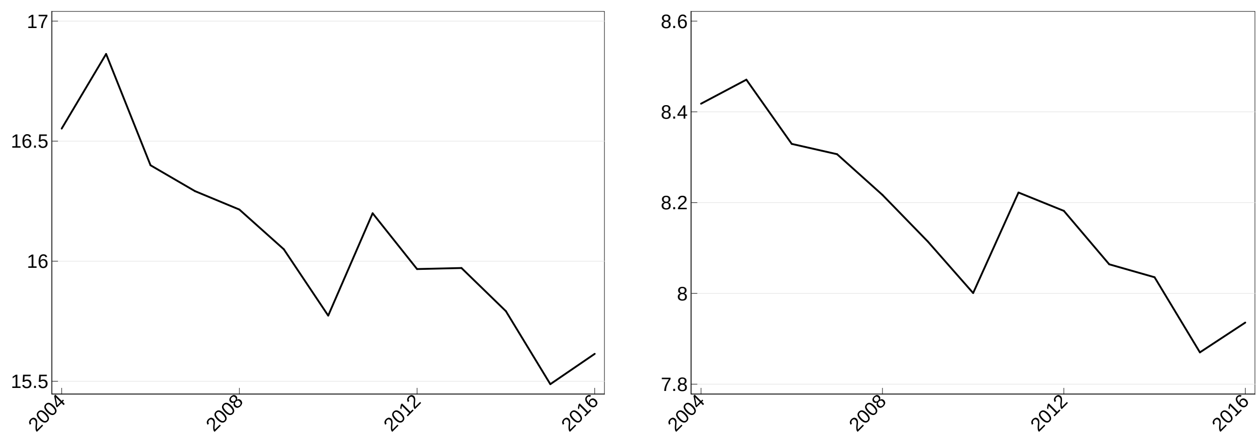
A.3 Measuring Varieties

A.3.1 Measuring Varieties Consumed Per Household

Figure A3 shows the average number of UPCs purchased per household constructed using the same weights as were used in equations (3) and (5).[6] 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 A3: Ave # Products Purchased by Each Household

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



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.[7]

[6]Since 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/(# households).

[7]These 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 dampened by around half.

A.3.2 Measuring Varieties Consumed in the Aggregate

The variety statistics thus far focus on the number of products purchased by individual households. 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 U.S. 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 A1 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%.[8] 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 be usable as inputs to our model or resulting welfare inference.

Table A1: 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	$\mathcal{H}^{HH}_{2004}$	$\mathcal{H}^{HH}_{2016}$	$\mathcal{H}^{HH}$ % change	$\mathcal{H}^{Agg}_{2004}$	$\mathcal{H}^{Agg}_{2016}$	$\mathcal{H}^{Agg}$ % 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 A1 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

[8]Note that for all of the calculations in Table A1, 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.

Table A2: Effect of Products with Small Aggregate Spending on Model Implied Annual Growth Rates

Agg Spend Threshold	j^* growth	$\tilde{N}$ growth	η growth	Utility growth from N	Utility growth from N, F, σ
\$0	2.0%	4.5%	-0.18%	0.56%	0.45%
\$25	2.0%	4.6%	-0.21%	0.57%	0.46%
\$50	2.1%	4.6%	-0.21%	0.57%	0.47%
\$250	2.1%	4.5%	-0.26%	0.56%	0.47%

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 annual growth in consumed varieties (j^*) is always 2-2.1%, 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 of UPCs in Nielsen with negligible spending.

A.4 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 *et al.* (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 *et al.* (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").

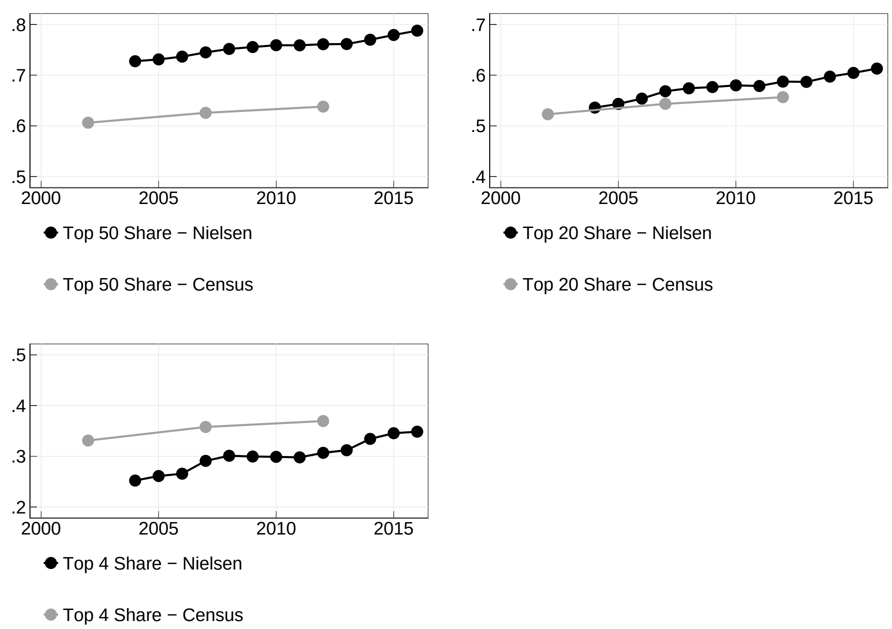
Finally, it is important to note that our sample covers the period 2004-2016 while census data starts in 1982 and is last available in 2012. The exact timing of concentration trends in Census data varies substantially, with many sectors exhibiting increases primarily in the period prior to our sample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

[9] It 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.

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 "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 A4 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 A4: 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.

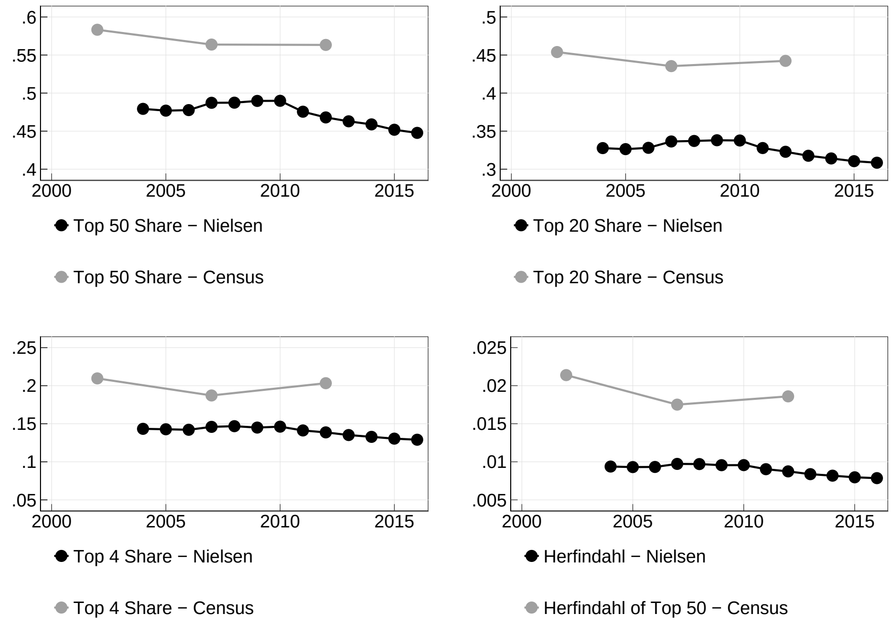
Second, UPCs for store-brand products map to the retailer rather than the actual manufacturer of the product. For example, Costco's "Kirkland" store-brand barcodes all map to "Costco", even though Costco does not actually produce most of these products. Although sometimes the actual producer can be identified (for example Kirkland Coffees are advertised as being roasted by Starbucks), this information is typically a trade-secret.[10] 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 A5 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.

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.

[10]Note that this specific anecdotal example is not drawn from the actual Nielsen micro data. License agreements prevent any disclosure of information about specific retailers or individual stores from this data.

Figure A5: Manufacturer Concentration



A.5 Additional Corroboration of Model Fit at Household Level

Figure A6 shows a number of additional rank comparisons of our main comparison of the household model fit discussed in Figure 6.

An alternative to the test of household-level model fit that we present in Figure 6 in the main text is to use the testable prediction from equation (14) that $\mathcal{H}^{HH}_{i,c}$ is proportional to $1/|\Omega_{i,c}|$. Indeed, when we pool categories, years, and households and regress $\ln|\Omega_{i,c}|$ on $-\ln\mathcal{H}^{\mathrm{HH}}_{i,c}$ (with category-year fixed effects), we get a coefficient of 0.89, which is close to the model-consistent value of 1, and a large R^2 of 0.82. The upper left panel of Figure A7 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.[11] In the upper right panel, we estimate these regressions separately for each category in 2016 and plot a histogram of the estimated slopes. The values are largely clustered around the model-consistent value of 1.

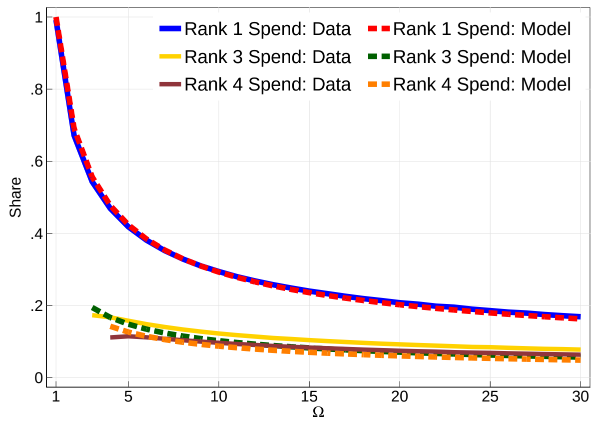
Next, rather than estimating the slope, we constrain it to equal 1 and back out the η_c values implied for each category. The model imposes the restriction that $0 < \eta_c < 1$ and the bottom left panel of Figure A7 shows that this restriction is satisfied in every category. The values of η_c range from lows of 0.08 (Baby Food) and 0.10 (Carbonated Beverages) to highs of 0.69 (Greeting Cards) and 0.97 (Yeast).[12]

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

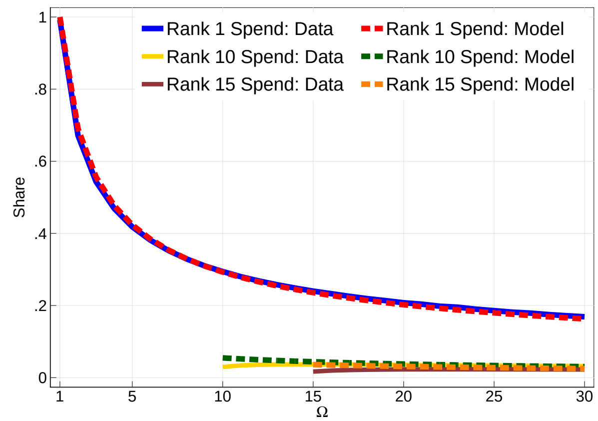
[12]The value of 0.08 for baby food implies that the typical household in this category has spending which is almost 4 times

Figure A6: Alternative Concentration Measures

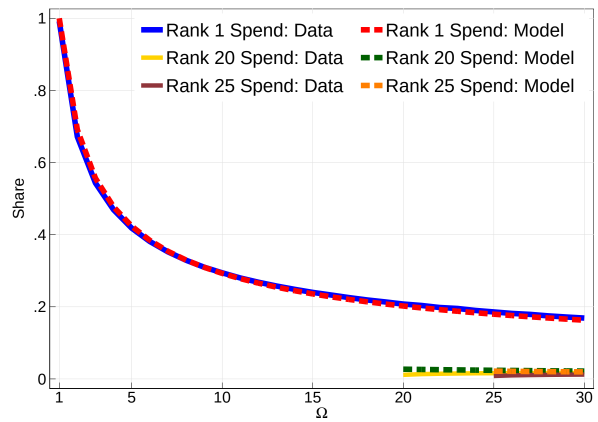
(a): Model vs Data: Additional Ranks 3 and 4



(b): Model vs Data: Additional Ranks 10 and 15



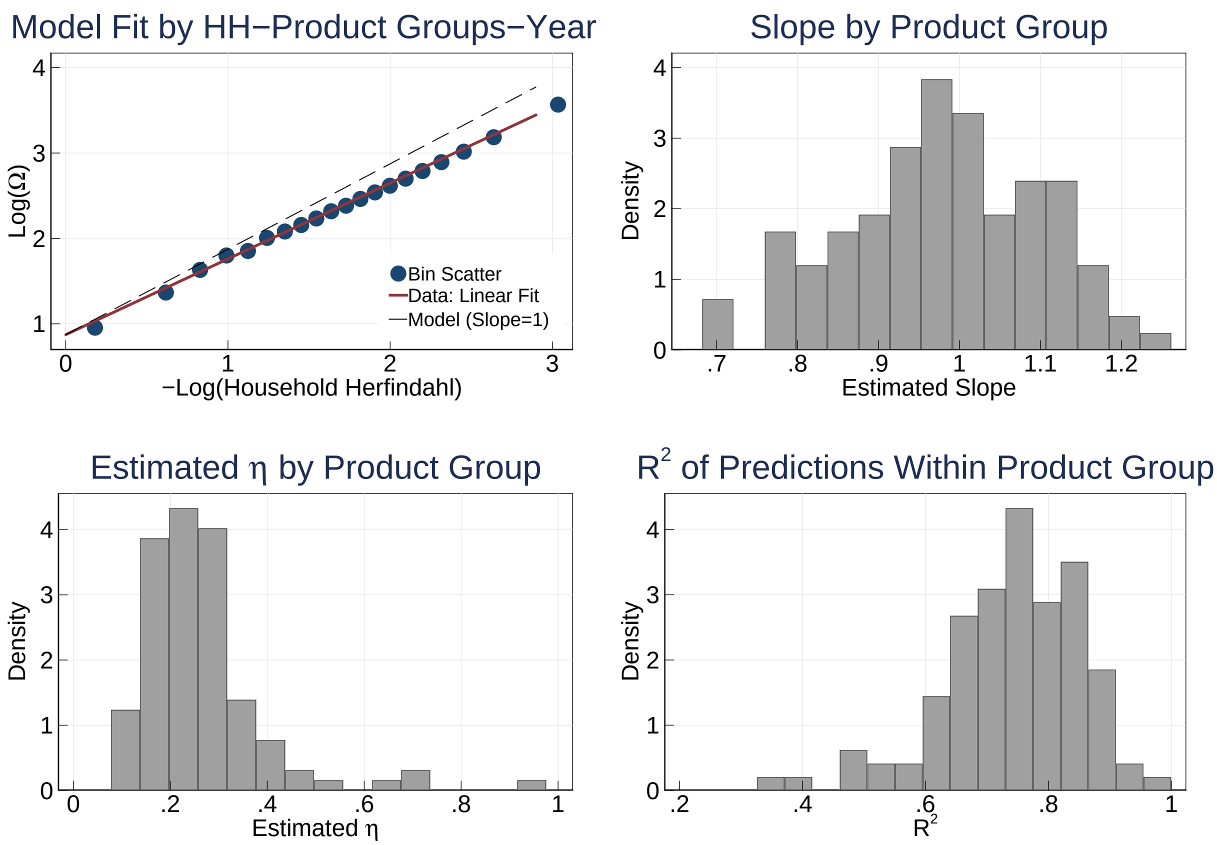
(c): Model vs Data: Additional Ranks 20 and 25



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.

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. $\theta \to \infty$) – the setup in many standard models – we cannot capture this large extent of sectoral heterogeneity in concentration.

Figure A7: Model Fit on Household-Category Data



A.6 Additional Empirical Results

Figure A8: Concentration Trends: Excluding Generics

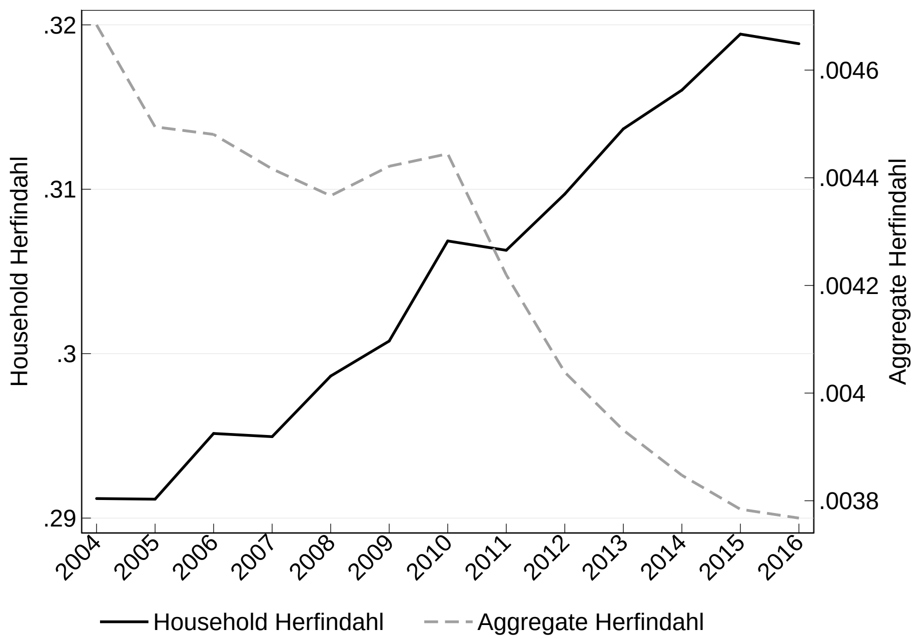


Figure A9: Concentration Trends: Including Category Composition Changes

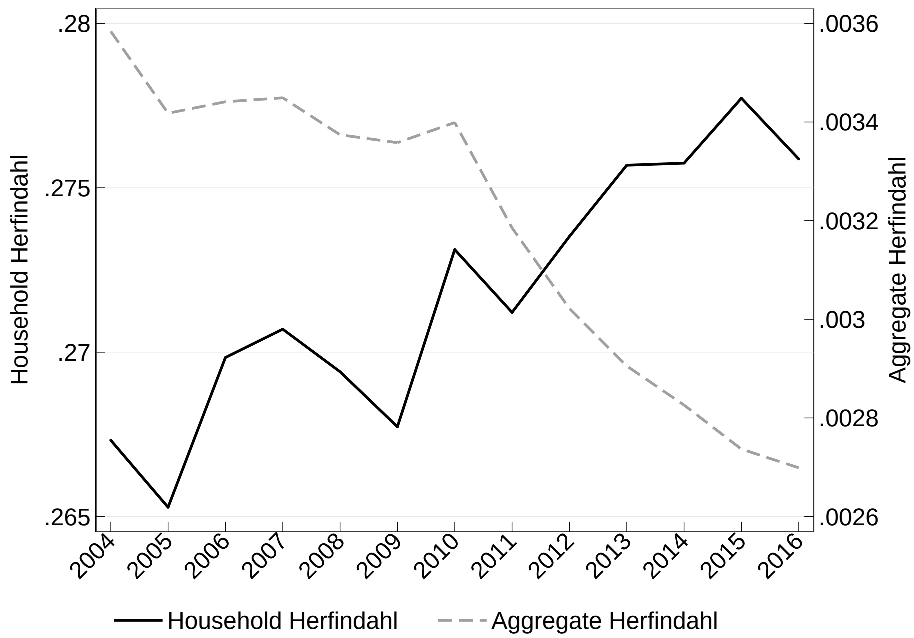


Figure A10: Concentration Trends: Brand Instead of UPC

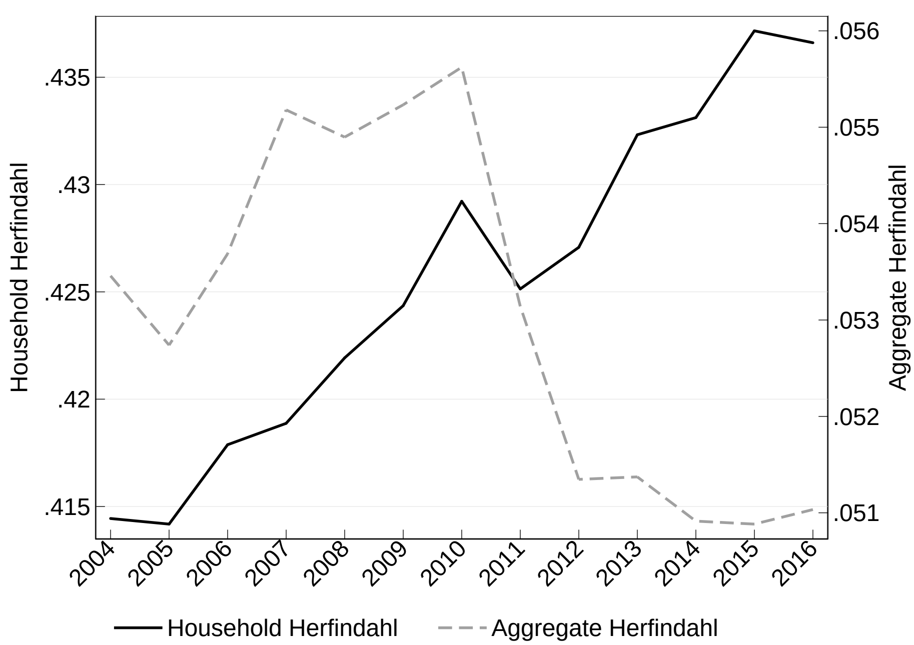


Figure A11: Concentration Trends: Product Module instead of Group

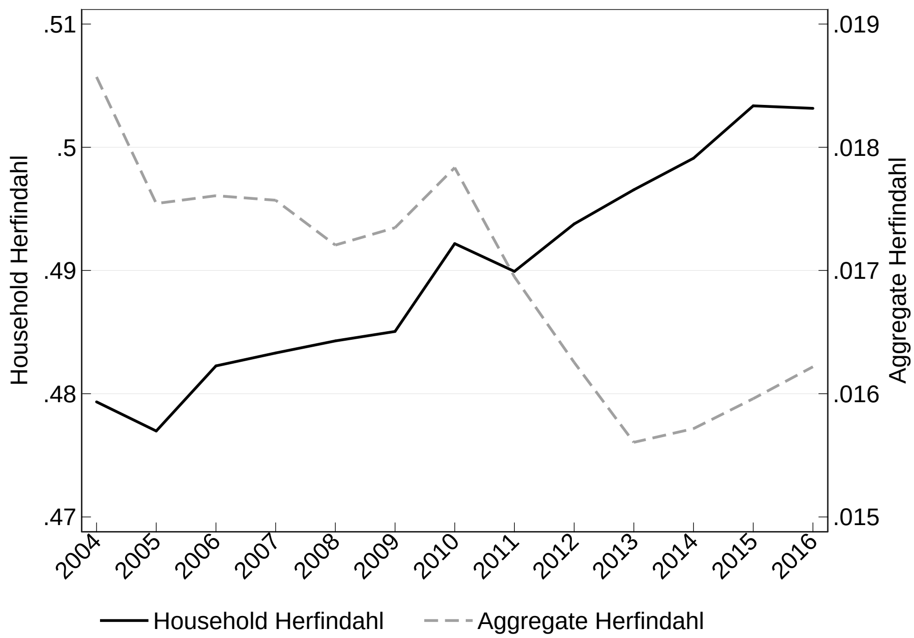
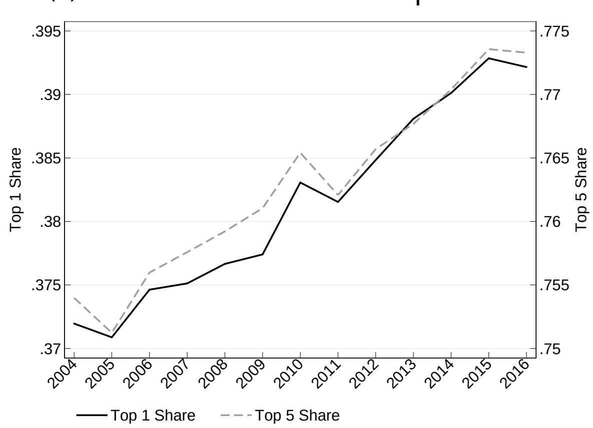


Figure A12: Alternative Concentration Measures

(a): Household Shares on Top Products



(b): Aggregate Shares on Top Products

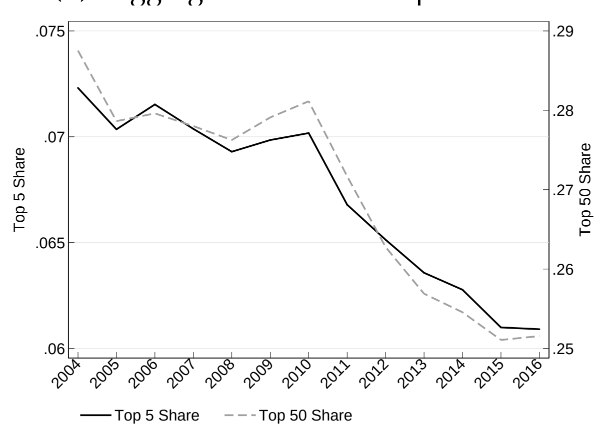
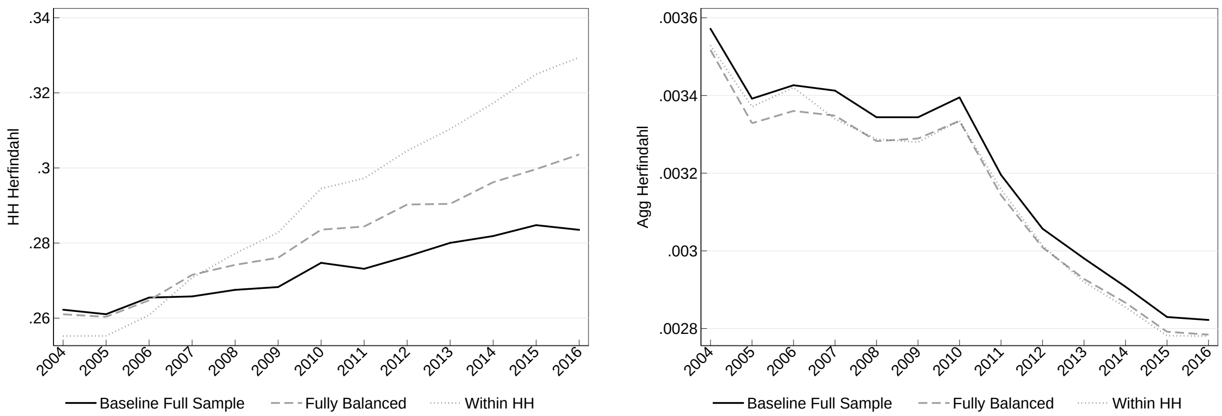


Figure A13: Concentration Trends for Different Samples

(a): Household Herfindahl **(b):** Aggregate Herfindahl



This figure recomputes Figure 1 for the baseline sample, a fully balanced sample and using only within household variation. The within-household variation specification calculates changes in concentration within household between years, averages these across household-categories and cumulates these changes over time beginning from the baseline level. The baseline sample averages individual *levels* instead of individual *changes* in household concentration. This means that it is more representative since it includes households only in the Nielsen panel for a single year, but it can change as households enter and leave the panel. Figure A2 shows that the baseline sample is a better fit to aggregate spending trends.

Figure A14: 2004-2016 Concentration Growth by Household Size

(a): Household Concentration **(b):** Aggregate Concentration

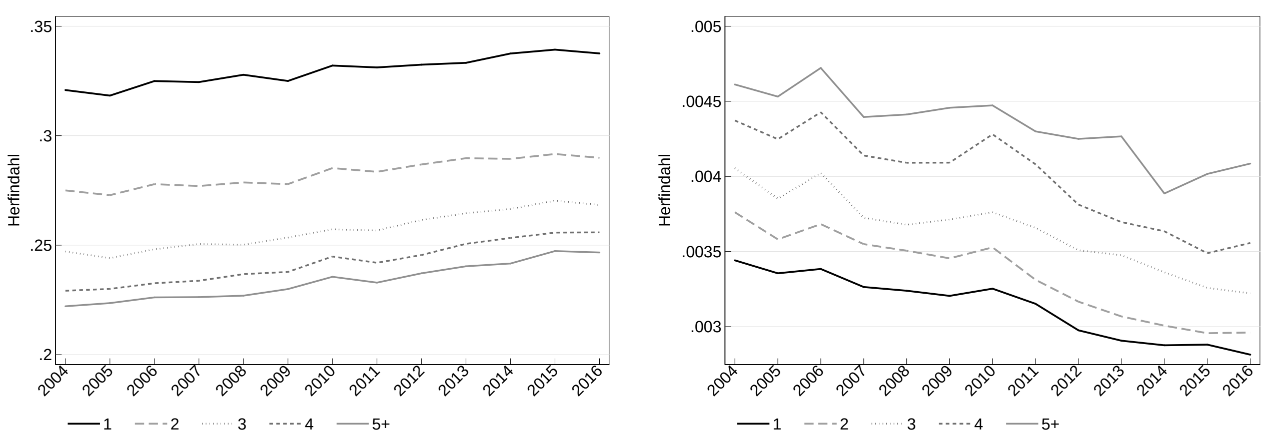
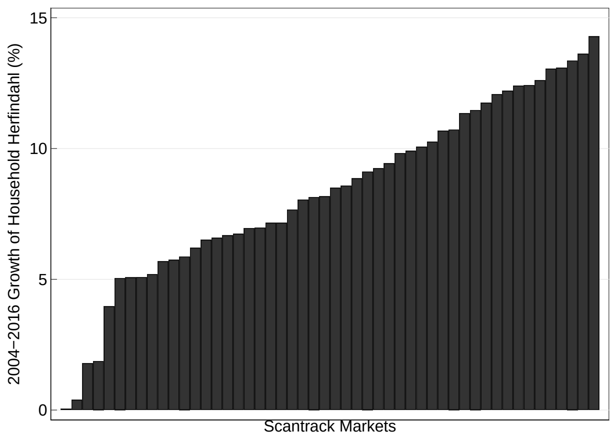


Figure A15: 2004-2016 Concentration Growth Within Location

(a): Household Concentration



(b): Aggregate Concentration

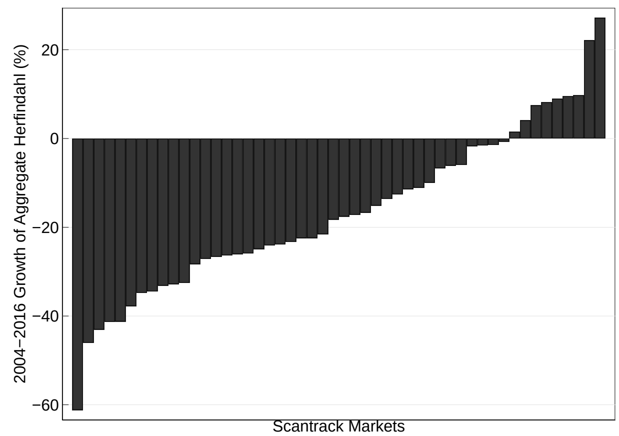
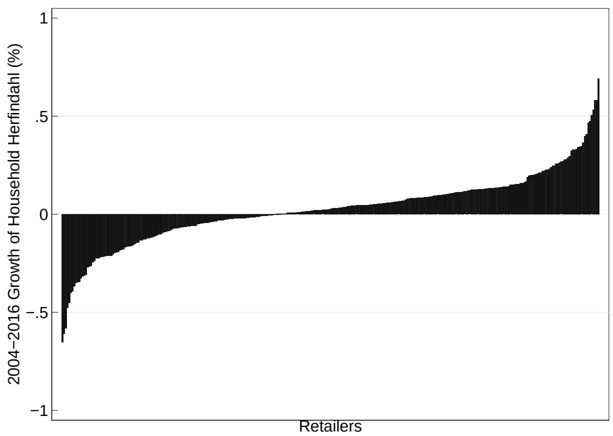


Figure A16: 2004-2016 Concentration Growth Within Retailer

(a): Household Concentration



(b): Aggregate Concentration

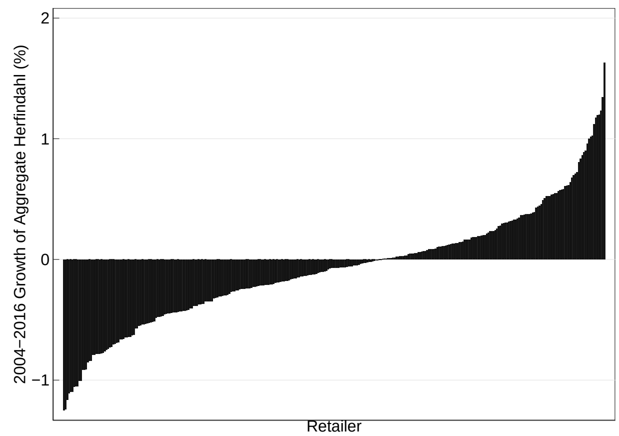
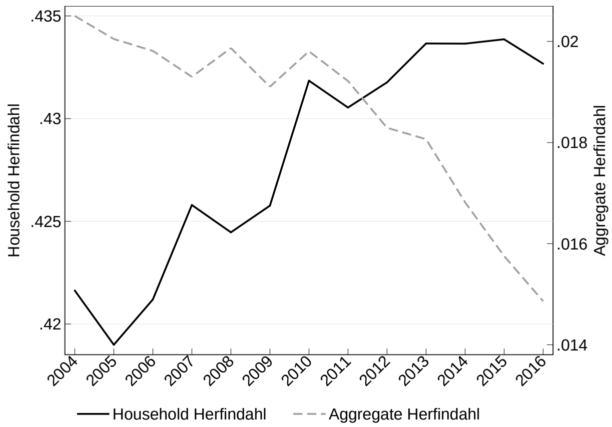
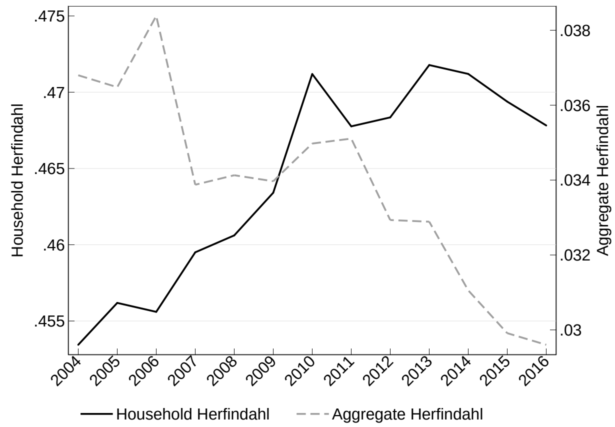


Figure A17: Within-Retailer Concentration Trends

(a): Include all Households

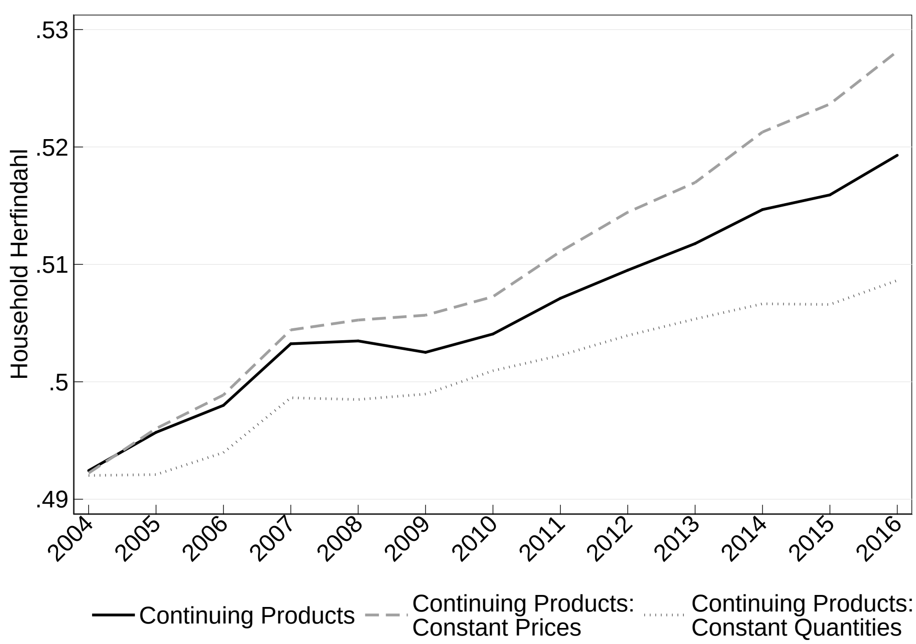


(b): Include Households Who Only Shop in One Retailer



This figure recomputes Figure 1 but defining market shares within retailer-categories instead of within categories. That is, we calculate household and aggregate concentration within individual retailer-category pairs and then average across all retailers and categories. The left panel includes all households while the right uses category spending only of households who shop in a single retailer for a given category to eliminate any composition effects from shifting expenditure across retailers within categories.

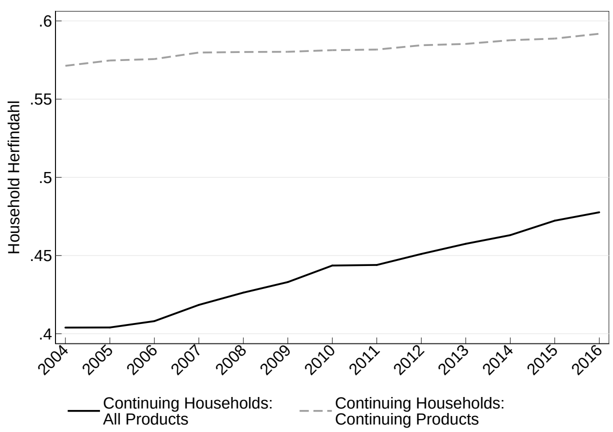
Figure A18: Intensive Margin P v. Q effects for UPCs



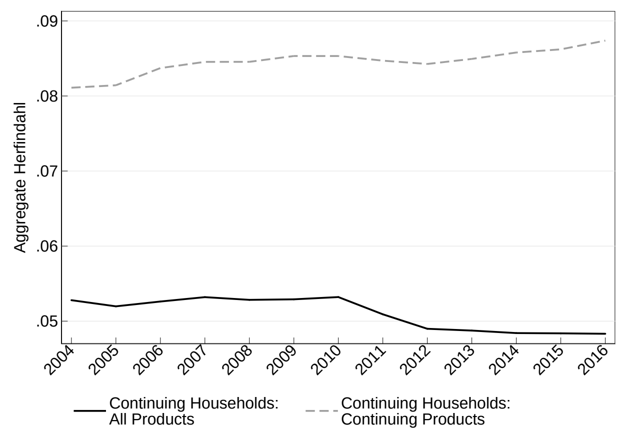
This figure recomputes the continuing product line in figure 5 but now either holding prices or holding quantities constant between t and $t+1$ for each household.

Figure A19: 2004-2016 Concentration growth for continuing vs. all products (brands)

(a): Household Herfindahl



(b): Aggregate Herfindahl



Appendix B. Theory Appendix

We start this section of the appendix with Subsection B.1, which provides detailed derivations of the main expressions presented in the body of the paper. Subsection B.2 offers additional results related to the simulation of our model. Finally, Subsection B.3 presents a version of our model where demand is assumed to be linear, rather than exhibiting a constant elasticity of substitution.

B.1 Derivation of Key Model Equations

In this appendix subsection, we provide detailed derivations of our expression for the price index P, the Household Herfindahl $\mathcal{H}^{\text{HH}}$, the CDF characterizing rank-values $R(r)$, the aggregate market share of a product s_j, the Aggregate Herfindahl $\mathcal{H}^{\text{Agg}}$, and elasticities of demand ε_j and markups at the product μ_j and aggregate levels μ^{Agg}.

B.1.1 Deriving the Price Index in Equation (10)

We start by showing how we can write $U_i = E/P - F\left(|\Omega_i|\right)^{\epsilon}$, where P is defined as in equation (10). The Lagrangian for the household consumption problem can be written as:

$$\mathcal{L} = \left(\int_{k\in\Omega_i} \left(\gamma_{i,k} C_{i,k}\right)^{\frac{\sigma-1}{\sigma}} dk\right)^{\frac{\sigma}{\sigma-1}} - F \times \left(|\Omega_i|\right)^{\epsilon} - \lambda\left(\int_{k\in\Omega_i} p_k C_{i,k} dk - E\right),$$

where λ is the multiplier on the budget constraint. Taking $|\Omega_i|$ as given and differentiating with respect to $C_{i,k}$ and setting equal to zero, we get:

$$\gamma_{i,k} C_{i,k} = \left(\lambda \frac{p_k}{\gamma_{i,k}} \frac{\sigma}{\sigma-1}\right)^{-\sigma},$$

for each good $k \in \Omega_i$. Taking the ratio of this first-order condition for good k to some other good j, rearranging, and integrating over goods k, we get:

$$E = \int_{k\in\Omega_i} C_{i,k} p_k dk = C_{i,j}\left(p_j\right)^{\sigma}\left(\gamma_{i,j}\right)^{1-\sigma}\left(\int_{k\in\Omega_i}\left(p_k\right)^{(1-\sigma)}\left(\gamma_{i,k}\right)^{\sigma-1} dk\right).$$

Defining $P_i = \left(\int_{k\in\Omega_i}\left(\tilde{\gamma}_{i,k}\right)^{\sigma-1} dk\right)^{\frac{1}{1-\sigma}}$, we can then re-write this expression as:

$$\gamma_{i,j} C_{i,j} = E\left(P_i\right)^{\sigma-1}\left(\tilde{\gamma}_{i,j}\right)^{\sigma}.$$

Next, we can raise both sides to the power $(\sigma-1)/\sigma$ and integrate over goods j to show that:

$$\int_{j\in\Omega_i}\left(\gamma_{i,j} C_{i,j}\right)^{\frac{\sigma-1}{\sigma}} dj = (E)^{\frac{\sigma-1}{\sigma}}\left(P_i\right)^{\frac{(\sigma-1)^2}{\sigma}}\int_{j\in\Omega_i}\left(\tilde{\gamma}_{i,j}\right)^{\sigma-1} dj = (E)^{\frac{\sigma-1}{\sigma}}\left(P_i\right)^{\frac{(\sigma-1)^2}{\sigma}}\left(P_i\right)^{1-\sigma}.$$

Finally, raising both sides to the $\sigma/\left(\sigma-1\right)$, gives:

$$\left(\int_{k\in\Omega_i}\left(\gamma_{i,k}C_{i,k}\right)^{\frac{\sigma-1}{\sigma}}dk\right)^{\frac{\sigma}{\sigma-1}}=E/P_i,$$

which will hold for any distribution of $\tilde{\gamma}_{i,k}$, regardless of the combinations of p_k and $\gamma_{i,k}$ that give rise to it.

Next, note that under the assumption that $G(y)$ is a Pareto distribution, the cutoff good is characterized by:

$$\frac{|\Omega_i|}{N}=\left(\frac{b}{\tilde{\gamma}_i^*}\right)^{\theta},$$

which implies:

$$\tilde{\gamma}_i^*=N^{\frac{1}{\theta}}\left(|\Omega_i|\right)^{-\frac{1}{\theta}}b. \tag{A1}$$

From the definition of P_i, we therefore have:

$$\begin{aligned}P=P_i &= \left(\int_{k\in\Omega_i}\left(\tilde{\gamma}_{i,k}\right)^{\sigma-1}dk\right)^{\frac{1}{1-\sigma}}=\left(N\int_{\tilde{\gamma}_i^*}^{\infty}y^{\sigma-1}dG(y)\right)^{\frac{1}{1-\sigma}}\\ &= N^{\frac{1}{1-\sigma}}\left(\frac{1}{\sigma-1-\theta}y^{\sigma-1-\theta}\Big|_{\tilde{\gamma}_i^*}^{\infty}\right)^{\frac{1}{1-\sigma}}\theta^{\frac{1}{1-\sigma}}b^{\frac{\theta}{1-\sigma}}=\left(1+\frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}}b^{\frac{\theta}{1-\sigma}}N^{\frac{1}{1-\sigma}}\left(\tilde{\gamma}_i^*\right)^{\frac{\theta}{\sigma-1}-1}.\end{aligned}$$

Substituting the value for $\tilde{\gamma}_i^*$ from equation (A1), we get:

$$P=P_i=\underbrace{\left(1+\frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}}b^{-1}}_{\text{Average Price}}\times\underbrace{\left(|\Omega_i|\right)^{\frac{1}{1-\sigma}}}_{\text{Variety Effects}}\times\underbrace{\left(\frac{|\Omega_i|}{N}\right)^{\frac{1}{\theta}}}_{\text{Selection Effects}}.$$

B.1.2 Deriving the Household Herfindahl $\mathcal{H}^{\text{HH}}$ in Equation (14)

We have:

$$\mathcal{H}^{\text{HH}}=\mathcal{H}_i^{\text{HH}}=N\int_{\tilde{\gamma}_i^*}^{\infty}\left(P_iy\right)^{2(\sigma-1)}G(y)dy=P^{2(\sigma-1)}\frac{N\theta b^{\theta}}{\theta-2\left(\sigma-1\right)}\left(\tilde{\gamma}^*\right)^{2(\sigma-1)-\theta}.$$

Substituting in the definition of $\tilde{\gamma}_i^*$ from equation (A1), we have:

$$\mathcal{H}^{\text{HH}}=P^{2(\sigma-1)}\frac{\theta}{\theta-2\left(\sigma-1\right)}N^{\frac{2(\sigma-1)}{\theta}}|\Omega|^{1-\frac{2(\sigma-1)}{\theta}}b^{2(\sigma-1)},$$

and substituting in the definition of P from equation (10), we get:

$$\mathcal{H}^{\text{HH}}=\left(1+\frac{1-\sigma}{\theta}\right)^2\frac{1}{1-\frac{2(\sigma-1)}{\theta}}\frac{1}{|\Omega|}.$$

Defining $\eta = 1 - \frac{2(\sigma-1)}{\theta}$, we then have:

$$\begin{aligned}
\mathcal{H}^{\mathrm{HH}} &= \frac{1}{\eta}\left(1+\frac{1-\sigma}{\theta}\right)^2 \frac{1}{|\Omega|} = \frac{1}{\eta}\left(1-\frac{2\left(\sigma-1\right)}{\theta}+\frac{\sigma-1}{\theta}\right)^2 \frac{1}{|\Omega|} \\
&= \frac{1}{\eta}\left(\eta+\frac{\sigma-1}{\theta}\right)^2 \frac{1}{|\Omega|} = \frac{\left(\eta+1\right)^2}{4\eta}\frac{1}{|\Omega|}.
\end{aligned}$$

B.1.3 Deriving the CDF $R(r)$ in Equation (17)

The goal here is to find the CDF of the rank value $r_{i,j} = (1-\alpha)\,j + \alpha x_{i,j}$ when these values are pooled across households i and products j. As a first step, let's pool only across households and solve for the conditional CDF for each product $j \in (0, N]$:

$$R_j(r) = Pr\left[(1-\alpha)\,j + \alpha x_{i,j} \leq r\right] = Pr\left[x_{i,j} < \frac{r-(1-\alpha)\,j}{\alpha}\right].$$

This yields:

$$R_j(r) = \begin{cases} 0, & 0 \leq r < (1-\alpha)\,j \\ \dfrac{r-(1-\alpha)\,j}{\alpha N}, & (1-\alpha)\,j \leq r < (1-\alpha)\,j + \alpha N \\ 1, & (1-\alpha)\,j + \alpha N \leq r \leq N. \end{cases}$$

We then can get the unconditional CDF by integrating these conditional CDFs across all products: $R(r) = \frac{1}{N}\int_0^N R_j(r)dj$. As an intermediate step, it will be useful to assume that $\alpha < 0.5$, which implies that $\alpha < 1-\alpha$, and to re-write the boundaries on the parameter space that define the three regions of the conditional CDF as follows:

$$R_j(r) = \begin{cases} 0, & min\left(\dfrac{r}{1-\alpha}, N\right) \leq j < N \\ \dfrac{r-(1-\alpha)\,j}{\alpha N}, & max\left(0, \dfrac{r-\alpha N}{1-\alpha}\right) \leq j < min\left(N, \dfrac{r}{1-\alpha}\right) \\ 1, & 0 \leq j \leq max\left(0, \dfrac{r-\alpha N}{1-\alpha}\right), \end{cases}$$

where the *min* and *max* conditions come from the restriction that $j \in (0, N]$.

We can then start by calculating the CDF in the first region – where $0 \leq r < \alpha N$ – as:

$$\begin{aligned}
R(r) &= \frac{1}{N}\int_0^N R_j(r)dj = \frac{1}{N}\int_0^0 1 \times dj + \frac{1}{N}\int_0^{\frac{r}{1-\alpha}} \frac{r-(1-\alpha)\,j}{\alpha N}dj + \frac{1}{N}\int_{\frac{r}{1-\alpha}}^N 0 \times dj \\
&= \frac{1}{N}\left(\frac{r}{\alpha N}\frac{r}{1-\alpha} - \frac{1-\alpha}{\alpha N}\frac{1}{2}\left(\frac{r}{1-\alpha}\right)^2\right) = \frac{1}{2}\left(\frac{r}{N}\right)^2 \frac{1}{\alpha\left(1-\alpha\right)},
\end{aligned}$$

where limits on the definite integrals come from evaluting the *min* and *max* operators above inside the region $0 \leq r < \alpha N$.

Next, we can calculate the CDF in the second region – where $\alpha N \leq r < (1-\alpha) N$ – as:

$$
\begin{aligned}
R(r) &= \frac{1}{N}\int_0^{\frac{r-\alpha N}{1-\alpha}} 1 \times dj + \frac{1}{N}\int_{\frac{r-\alpha N}{1-\alpha}}^{\frac{r}{1-\alpha}} \frac{r-(1-\alpha)j}{\alpha N} dj + \frac{1}{N}\int_{\frac{r}{1-\alpha}}^{N} 0 \times dj \\
&= \frac{1}{N}\frac{r-\alpha N}{1-\alpha} + \frac{1}{N}\left(\frac{r}{\alpha N}\frac{r}{1-\alpha} - \frac{(1-\alpha)}{\alpha N}\frac{1}{2}\left(\frac{r}{1-\alpha}\right)^2\right) - \frac{1}{N}\left(\frac{r}{\alpha N}\frac{r-\alpha N}{1-\alpha} - \frac{(1-\alpha)}{\alpha N}\frac{1}{2}\left(\frac{r-\alpha N}{1-\alpha}\right)^2\right) \\
&= \frac{1}{N}\left(r^2 2\alpha N(1-\alpha)\right) + \frac{r}{(1-\alpha)N} - \frac{\alpha}{1-\alpha} - \frac{1}{2N}\left(\frac{r^2}{\alpha N(1-\alpha)} - \frac{\alpha N}{1-\alpha}\right) \\
&= \frac{r}{N}\frac{1}{1-\alpha} - \frac{1}{2}\frac{\alpha}{1-\alpha}.
\end{aligned}
$$

Finally, we can calculate the CDF in the third region – where $(1-\alpha) N \leq r \leq N$ – as:

$$
\begin{aligned}
R(r) &= \frac{1}{N}\int_0^{\frac{r-\alpha N}{1-\alpha}} 1 \times dj + \frac{1}{N}\int_{\frac{r-\alpha N}{1-\alpha}}^{N} \frac{r-(1-\alpha)j}{\alpha N} dj + \frac{1}{N}\int_{N}^{N} 0 \times dj \\
&= \frac{1}{N}\frac{r-\alpha N}{1-\alpha} + \frac{1}{N}\left(\frac{r}{\alpha} - \frac{N}{2}\frac{1-\alpha}{\alpha} - \frac{1}{2}\frac{r^2-\alpha^2 N^2}{\alpha(1-\alpha)N}\right) \\
&= \frac{1}{2}\left(\frac{r}{N}\right)^2 \frac{1}{\alpha(1-\alpha)} + \frac{r}{N}\frac{1}{\alpha(1-\alpha)} - \frac{1}{2}\left(\frac{1-\alpha}{\alpha} + \frac{\alpha}{1-\alpha}\right).
\end{aligned}
$$

Collecting these results yields the CDF for r in Equations (17).

B.1.4 Deriving the Aggregate Market Share s_j in Equation (22)

To derive the aggregate market share for a product j, we need to integrate spending shares across all households that buy j, where their spending shares are heterogeneous due to their receipt of an idiosyncratic taste shock $x_{i,j}$. Toward that end, we will use four new expressions:

1. Noting that $\tilde{\gamma}_{i,j} = G^{-1}\left(1 - R(r_{i,j})\right) = b\left(R(r_{i,j})\right)^{-\frac{1}{\theta}}$, we can substitute in to write the household's spending share on good j as:

   $$s_{i,j} = P_i^{\sigma-1}\tilde{\gamma}_{i,j}^{\sigma-1} = (P_i b)^{\sigma-1}\left(R\left(r_{i,j}\right)\right)^{-\frac{\sigma-1}{\theta}},$$

   if $R(r_{i,j}) \leq |\Omega|/N$, and zero otherwise.

2. We need to solve for the cutoff rank value r^* so we can determine the worst idiosyncratic draw x_j^* for product j that still yields positive spending on that variety. *Focusing only on the first of the three regions* from the CDFs above, we have that the cutoff r^* satisfies $R(r^*) = |\Omega|/N$ or:

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

   Solving the resulting quadratic equation for a positive root in $(0, N]$ leaves: $r^* = \left(2(1-\alpha)\alpha|\Omega|N\right)^{\frac{1}{2}}$.

3. We can then solve for the highest j good that experiences any positive consumption in the economy, j^*, as its rank value will equal r^* even when the idiosyncratic draw is the best possible case of $x = 0$. It will satisfy $r^* = (1-\alpha) j^*$, or $j^* = r^* / (1-\alpha) = \left(\frac{2\alpha|\Omega|N}{1-\alpha}\right)^{\frac{1}{2}}$.

4. Note that for a given good j, the worst possible idiosyncratic taste draw, x_j^* that yields positive consumption of j satisfies: $(1-\alpha) j + \alpha x_j^* = r^* = (1-\alpha) j^*$, or $x_j^* = \frac{1-\alpha}{\alpha}(j^* - j)$.

We can then solve for the aggregate expenditure share of good j as:

$$
\begin{aligned}
s_j &= \frac{1}{\int_i Edi}\int_i Es_{i,j}di = \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\int_0^{x_j^*}\left(R\left((1-\alpha)j+\alpha x\right)\right)^{\frac{\eta-1}{2}}\frac{dx}{N} \\
&= \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\int_0^{\frac{1-\alpha}{\alpha}(j^*-j)}\left(R\left((1-\alpha)j+\alpha x\right)\right)^{\frac{\eta-1}{2}}\frac{dx}{N} \\
&= \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\int_0^{\frac{1-\alpha}{\alpha}(j^*-j)}\left(\frac{1}{2}\left(\frac{(1-\alpha)j+\alpha x}{N}\right)^2\frac{1}{\alpha(1-\alpha)}\right)^{\frac{\eta-1}{2}}\frac{dx}{N} \\
&= \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\int_0^{\frac{1-\alpha}{\alpha}(j^*-j)}2^{\frac{1-\eta}{2}}N^{1-\eta}\left(\alpha(1-\alpha)\right)^{\frac{1-\eta}{2}}\left((1-\alpha)j+\alpha x\right)^{\eta-1}\frac{dx}{N} \\
&= (\eta+1)\left(2N|\Omega|\right)^{-\frac{\eta+1}{2}}\left(\alpha(1-\alpha)\right)^{\frac{1-\eta}{2}}\int_0^{\frac{1-\alpha}{\alpha}(j^*-j)}\left((1-\alpha)j+\alpha x\right)^{\eta-1}dx \\
&= (\eta+1)\left(2N|\Omega|\right)^{-\frac{\eta+1}{2}}\left(\alpha(1-\alpha)\right)^{\frac{1-\eta}{2}}\frac{1}{\alpha\eta}\left((1-\alpha)j+\alpha x\right)^{\eta}\Big|_0^{\frac{1-\alpha}{\alpha}(j^*-j)} \\
&= \frac{\eta+1}{\eta}\left(\left(\frac{2\alpha N|\Omega|}{1-\alpha}\right)^{\frac{1}{2}}\right)^{-\frac{\eta+1}{2}}\left((j^*)^{\eta}-j^{e}ta\right) \\
&= \frac{\eta+1}{\eta j^*}\left(1-\left(\frac{j}{j^*}\right)^{\eta}\right).
\end{aligned}
$$

B.1.5 Deriving the Aggregate Herfindahl $\mathcal{H}^{\text{Agg}}$ in Equation (23)

The Aggregate Herfindahl is calculated as:

$$
\begin{aligned}
\mathcal{H}^{\text{Agg}} &= \int_0^{j^*}s_j^2dj = \left(\frac{\eta+1}{\eta j^*}\right)^2\int_0^{j^*}\left(1-\left(\frac{j}{j^*}\right)^{\eta}\right)^2dj \\
&= \left(\frac{\eta+1}{\eta j^*}\right)^2\int_0^{j^*}\left(1-2\left(\frac{j}{j^*}\right)^{\eta}+\left(\frac{j}{j^*}\right)^{2\eta}\right)dj \\
&= \left(\frac{\eta+1}{\eta j^*}\right)^2\left[j-2\left(\frac{1}{j^*}\right)^{\eta}\frac{j^{\eta+1}}{\eta+1}+\left(\frac{1}{j^*}\right)^{2\eta}\frac{j^{2\eta+1}}{2\eta+1}\right]_0^{j^*} \\
&= \left(\frac{\eta+1}{\eta j^*}\right)^2\left[j^*-2\frac{j^*}{\eta+1}+\frac{j^*}{2\eta+1}\right] \\
&= \frac{2(\eta+1)}{2\eta+1}\frac{1}{j^*} \\
&= \frac{2(\eta+1)}{2\eta+1}\left(\frac{1}{2|\Omega|\tilde{N}}\right)^{\frac{1}{2}},
\end{aligned}
$$

where we define $\tilde{N} = \alpha N / (1-\alpha)$.

B.1.6 Deriving Elasticities ε_j and Markups μ_j and μ^{Agg} in Equations (32), (33), and (34)

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:

$$s_j = \frac{1}{N}\int_0^{x_j^*} s_{i_x,j}dx, \tag{A2}$$

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. We take the partial derivative of s_j in equation (A2) with respect to p_j to get:

$$\frac{\partial s_j}{\partial p_j} = \frac{1}{N}\left(\int_0^{x_j^*} \frac{\partial s_{i_x,j}}{\partial p_j}dx + s_{i_{x_j^*},j}\frac{\partial x_j^*}{\partial p_j}\right), \tag{A3}$$

where the right hand side of equation (A3) follows from Leibniz's rule. The first term can be solved using equation (18) as:

$$\frac{\partial s_{i_x,j}}{\partial p_j} = \frac{\partial P^{\sigma-1}p_j^{1-\sigma}\gamma_{i,j}^{\sigma-1}}{\partial p_j} = (1-\sigma)\frac{s_{i_x,j}}{p_j}, \tag{A4}$$

where we take the aggregate price index P as fixed. Moving on to the second term, we can evaluate equation (18) at the marginal household with taste x_j^* to get:

$$s_{i_{x_j^*},j}\frac{\partial x_j^*}{\partial p_j} = \frac{\eta+1}{2}N^{\frac{\eta-1}{2}}|\Omega|^{-\frac{\eta+1}{2}}\left(R\left(r^*\right)\right)^{\frac{\eta-1}{2}}\frac{\partial x_j^*}{\partial p_j}. \tag{A5}$$

Substituting equations (A4) and (A5) back into equation (A3), we get:

$$\begin{aligned}
\frac{\partial s_j}{\partial p_j} &= (1-\sigma)\frac{1}{p_j}\frac{1}{N}\int_0^{x_j^*} s_{i_x,j}dx + \frac{1}{N}s_{i_{x_j^*},j}\frac{\partial x_j^*}{\partial p_j} \\
&= (1-\sigma)\frac{s_j}{p_j} + \frac{1}{N}\frac{\eta+1}{2}\left(|\Omega|\right)^{-\frac{1+\eta}{2}}N^{\frac{\eta-1}{2}}\left(R\left(r^*\right)\right)^{\frac{\eta-1}{2}}\frac{\partial x_j^*}{\partial p_j} \\
&= (1-\sigma)\frac{s_j}{p_j} + \frac{1}{N}\frac{\eta+1}{2}\left(|\Omega|\right)^{-\frac{1+\eta}{2}}N^{\frac{\eta-1}{2}}\left(\frac{1-\alpha}{2\alpha N^2}\left(j^*\right)\right)^{\frac{\eta-1}{2}}\frac{\partial x_j^*}{\partial p_j} \\
&= (1-\sigma)\frac{s_j}{p_j} + \frac{\eta+1}{2N|\Omega|}\frac{\partial x_j^*}{\partial p_j}.
\end{aligned} \tag{A6}$$

To approximate $\partial x_j^*/\partial p_j$, we start with the relationship:

$$\frac{1}{2}\left(\frac{r_{i,j}}{N}\right)^2\frac{1}{\alpha(1-\alpha)} = R\left((1-\alpha)j + \alpha x_{i,j}\right) = 1 - G\left(\frac{\gamma_{i,j}}{p_j}\right) = b^\theta\gamma_{i,j}^{-\theta}p_j^\theta, \tag{A7}$$

and differentiate to yield:

$$\frac{r_{i,j}}{1-\alpha}\frac{1}{N^2}\frac{\partial x_{i,j}}{\partial p_j} = \theta b^\theta \gamma_{i,j}^{-\theta} p_j^{\theta-1}, \tag{A8}$$

where we've substituted $\partial r_{i,j}/\partial x_{i,j} = \alpha$. We then evaluate equation (A8) at $r_{i,j} = r^*$ and $\tilde{\gamma}_{i,j} = \tilde{\gamma}^*$ using equation (19) 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:

$$\begin{aligned}
\frac{\partial x_j^*}{\partial p_j} &= -\theta b^\theta \gamma_{i,j}^{-\theta} p_j^{\theta-1} N^2 \frac{1-\alpha}{r^*} \\
&= -\theta b^\theta \left(\tilde{\gamma}^*\right)^{-\theta} p_j^{-1} N^2 \frac{1}{j^*} \\
&= -\theta \frac{|\Omega|}{N} p_j^{-1} N^2 \frac{1}{j^*} \\
&= -\theta |\Omega| N p_j^{-1} \frac{1}{j^*} \\
&= -\frac{\theta}{j^*} |\Omega| N \frac{1}{p_j}.
\end{aligned}$$

Inserting this into equation (A6), we have:

$$\begin{aligned}
\frac{\partial s_j}{\partial p_j} &= (1-\sigma)\frac{s_j}{p_j} + \frac{\eta+1}{2N|\Omega|}\frac{\partial x_j^*}{\partial p_j} \\
&= (1-\sigma)\frac{s_j}{p_j} - \frac{\eta+1}{2N|\Omega|}\frac{\theta}{j^*}|\Omega| N \frac{1}{p_j} \\
&= (1-\sigma)\frac{s_j}{p_j} - \frac{\eta+1}{j^*}\frac{\theta}{2}\frac{1}{p_j} \\
&= (1-\sigma)\frac{s_j}{p_j} - \frac{\eta\theta}{2\left(1-\left(\frac{j}{j^*}\right)^\eta\right)}\frac{s_j}{p_j},
\end{aligned} \tag{A9}$$

where the last substitution uses the definition of product share in equation (22). Equation (31) implies that product j's price elasticity of demand ε_j can be written as:

$$\varepsilon_j = 1 - \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = \underbrace{\sigma}_{\text{Intensive Margin}} + \underbrace{\left(1-\left(\frac{j}{j^*}\right)^\eta\right)^{-1}\left[\theta/2 - (\sigma-1)\right]}_{\text{Extensive Margin}} \quad > \quad \sigma. \tag{A10}$$

The markup μ_j then be written (in gross terms) as:

$$\mu_j = \frac{\varepsilon_j}{\varepsilon_j - 1} = \frac{\sigma + \frac{\theta(\eta+1)}{2j^* s_j}}{\sigma + \frac{\theta(\eta+1)}{2j^* s_j} - 1}. \tag{A11}$$

The aggregate markup μ^{Agg} is equal to the ratio of aggregate sales to aggregate costs. Using equations (22) and (A10), it can be written as:

$$\mu^{\text{Agg}} = \frac{\int_0^{j^*} s_j dj}{\int_0^{j^*} s_j \frac{\varepsilon_j - 1}{\varepsilon_j} dj} = \left[\frac{\theta + (\sigma-1)^2}{\sigma^2} - \frac{1}{2}\frac{\eta\theta^2}{\sigma^2}\left(\frac{\eta+1}{2+\theta}\right) \times {}_2F_1\left(1, \frac{1}{\eta}; 1 + \frac{1}{\eta}; \frac{2\sigma}{2+\theta}\right)\right]^{-1}, \tag{A12}$$

where we've used Mathematica to evaluate and simplify this final expression.

B.1.7 Testable Predictions at Household Level

We now use the model to derive a relationship between a product's share of a household's total spending and its rank in that household's bundle, given the total number of goods in the bundle. In particular, we show how to convert our model which assumes a continuum of products are purchased into a discrete interpretation which can be tested directly in the data.

We start with expression (13) from the paper, which states that for household i that ranks product $k \leq |\Omega|$, k's spending share will equal:

$$\begin{aligned} s_{i,k} &= (P\tilde{\gamma}_{i,k})^{\sigma-1} && \text{(A13)} \\ &= \left(\left(1 + \frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}} b^{-1} (|\Omega_i|)^{\frac{1}{1-\sigma}} \left(\frac{|\Omega_i|}{N}\right)^{\frac{1}{\theta}} \tilde{\gamma}_{i,k}\right)^{\sigma-1} && \text{(A14)} \\ &= \left(1 + \frac{1-\sigma}{\theta}\right) (|\Omega_i|)^{-1} b^{1-\sigma} \left(\frac{|\Omega_i|}{N}\right)^{\frac{\sigma-1}{\theta}} (\tilde{\gamma}_{i,k})^{\sigma-1}, && \text{(A15)} \end{aligned}$$

where the second line substitutes in the definition of P from equation (10) in the paper.

Define the percentile $p_{i,k}$ of good k in i's bundle as:

$$p_{i,k} = \frac{1 - G(\tilde{\gamma})}{1 - G(\tilde{\gamma}^*)} = \frac{N}{|\Omega_i|}\left(\frac{\tilde{\gamma}_{i,k}}{b}\right)^{-\theta}, \tag{A16}$$

where a value of $p_{i,k} = 0.10$, for example, means variety k is at the 10th percentile out of all those that are consumed and is thus preferred to 90 percent of the goods consumed by that household. Substituting, this then allows us to write the expenditure share of a good as a decreasing function of its percentile (ranging from 0 to 1) in terms of how preferred it is among the goods that are consumed:

$$s_{i,k} = \frac{\eta+1}{2|\Omega_i|} p_{i,k}^{\frac{\eta-1}{2}}. \tag{A17}$$

Note that integrating over all consumed goods k, the shares sum to one:

$$\int_0^1 \frac{\eta+1}{2|\Omega_i|} p_{i,k}^{\frac{\eta-1}{2}} |\Omega_i| dp_{i,k} = 1. \tag{A18}$$

However, our data is discrete. One might be tempted to simply substitute $p_{i,k} = 1/|\Omega_i|$ for the highest share good in our data, and $p_{i,k} = 2/|\Omega_i|$ for the second, and so-on, and then estimate the equation:

$$\ln(s_{i,k}) = \ln\left(\frac{\eta+1}{2|\Omega_i|}\right) + \frac{\eta-1}{2}\ln(p_{i,k}) + \epsilon_{i,k}, \tag{A19}$$

constraining the slope to be $|\Omega_i|$ times the intercept and then comparing data relative to model predictions for various values of Ω_i. The problem with this is that since our data is discrete, the equation does not exactly hold. For example, if there are 8 goods, using the formula with $p = \{1/8, 2/8, ..., 1\}$, does not imply shares that sum to one.

We therefore derive an analog of equation (A17) for a world with discrete goods. In particular, define $s_{i,K}^{|\Omega_i|}$ for $K \in \{1, 2, ..., |\Omega_i|\}$ as the combined spending share of all varieties in the span of $[(K-1)/|\Omega_i|, K/|\Omega_i|]$. We then associate $s_{i,K}^{|\Omega_i|}$ with the spending share of good with rank K in the data. We calculate:

$$\begin{aligned} s_{i,K}^{|\Omega_i|} &= \int_{(K-1)/|\Omega_i|}^{K/|\Omega_i|} \frac{\eta+1}{2|\Omega_i|} p_{i,k}^{\frac{\eta-1}{2}} |\Omega_i| dp_{i,k} && \text{(A20)} \\ &= (|\Omega_i|)^{-\frac{\eta+1}{2}} \left[K^{\frac{\eta+1}{2}} - (K-1)^{\frac{\eta+1}{2}}\right]. && \text{(A21)} \end{aligned}$$

These shares then are functions of both $|\Omega_i|$ and η. For example, the simplest case is the share spent on the most preferred good (i.e. $K = 1$) which can be written as:

$$s_{i,1}^{|\Omega_i|} = (|\Omega_i|)^{-\frac{\eta+1}{2}}. \tag{A22}$$

The other cases are similar but the bracketed term in equation (A21) does not simplify. Figure 6 in the paper then demonstrates that this theoretical relationship (A21) aligns closely with the spending patterns in our data.

B.1.8 Relationship to Feenstra (1994)

In this subsection, we connect the standard Feenstra (1994) correction with the price index in our environment. Let F_t denote the Feenstra correction at t, defined as:

$$F_t = \frac{P_t/P_{t-1}}{P_t^{\text{Conventional}}/P_{t-1}^{\text{Conventional}}}, \tag{A23}$$

where P_t is the exact price index at t corresponding to demand over varieties j with a constant elasticity of substitution (CES) and where $P_t^{\text{Conventional}}$ is a "conventional" price index that is measured entirely using varieties that are consumed in two consecutive periods:

$$\frac{P_t^{\text{Conventional}}}{P_{t-1}^{\text{Conventional}}} = \prod_j (p_{j,t}/p_{j,t-1})^{\omega_{j,t}}, \tag{A24}$$

where $p_{j,t}$ is the price of good j at t and the weights $\omega_{j,t}$ are Sato-Vartia weights and are functions of the expenditure shares of good j in t and $t-1$.

There are four key quantities and one key parameter in the Feenstra correction. It equals:

$$F_t = \left(\frac{F_{1,t}/F_{2,t}}{F_{3,t}/F_{4,t}} \right)^{-\frac{1}{\sigma-1}}. \tag{A25}$$

$F_{1,t}$ is the spending at t, $F_{2,t}$ is the spending at t on goods that were also purchased at $t-1$, $F_{3,t}$ is the spending at $t-1$, and $F_{4,t}$ is the spending at $t-1$ on goods that will eventually be purchased at t.

To express the Feenstra correction at the household level in our environment, consider an increase in N from N_{t-1} to N_t that increases Ω_{t-1} to Ω_t. Note that $N_t > N_{t-1}$ causes $|\Omega_t| > |\Omega_{t-1}|$ and $\tilde{\gamma}_t^* > \tilde{\gamma}_{t-1}^*$, because the increase in N exceeds the increase in $|\Omega|$ (i.e. $|\Omega|/N$ goes down). So in period t, all products with values of $\tilde{\gamma} \in \left(\tilde{\gamma}_{t-1}^*, \tilde{\gamma}_t^*\right)$ that were consumed in $t-1$ are dropped, and every value of $\tilde{\gamma}$ that was purchased in t was also purchased in $t-1$. However, since the density of varieties with any given $\tilde{\gamma}$ value is N_t/N_{t-1} higher, we have:

$$F_{2,t} = E\frac{N_{t-1}}{N_t}. \tag{A26}$$

For $F_{4.t}$, we simply want to calculate spending during $t-1$ on all $\tilde{\gamma}$ values greater than $\tilde{\gamma}_t^*$, since those varieties would also have been consumed at t:

$$\begin{aligned}
F_{4,t} &= EN_{t-1}\int_{\tilde{\gamma}_t^*}^{\infty} (P_{t-1})^{\sigma-1} (y)^{\sigma-1} dG(y) && \text{(A27)}\\
&= EN_{t-1} (P_{t-1})^{\sigma-1} \int_{\tilde{\gamma}_t^*}^{\infty} (y)^{\sigma-1} dG(y) && \text{(A28)}\\
&= EN_{t-1} (P_{t-1})^{\sigma-1} \left(\frac{\theta b^\theta}{\theta - \sigma + 1} \right) (\tilde{\gamma}_t^*)^{\sigma-1-\theta} && \text{(A29)}\\
&= Eb^{\theta+1-\sigma} \left(\frac{|\Omega_{t-1}|}{N_{t-1}} \right)^{\frac{\sigma-1-\theta}{\theta}} (\tilde{\gamma}_t^*)^{\sigma-1-\theta} && \text{(A30)}\\
&= E \left(\frac{|\Omega_t|/|\Omega_{t-1}|}{N_t/N_{t-1}} \right)^{\frac{\theta+1-\sigma}{\theta}} && \text{(A31)}
\end{aligned}$$

Remember that the terms F_1 and F_3 just capture spending in t and $t-1$, so they are equal: $F_{1,t} = F_{3,t} = E$.

Putting this all together into equation (A25), we have

$$F_t = \left(\frac{F_{4,t}}{F_{2,t}} \right)^{-\frac{1}{\sigma-1}} = \left(\frac{\left(\frac{|\Omega_t|/|\Omega_{t-1}|}{N_t/N_{t-1}} \right)^{\frac{\theta+1-\sigma}{\theta}}}{\frac{N_{t-1}}{N_t}} \right)^{-\frac{1}{\sigma-1}}$$

$$= \left(\frac{|\Omega_t|}{|\Omega_{t-1}|} \left(\frac{|\Omega_t|/|\Omega_{t-1}|}{N_t/N_{t-1}} \right)^{\frac{1-\sigma}{\theta}} \right)^{-\frac{1}{\sigma-1}}$$

$$= \underbrace{\left(\frac{|\Omega_t|}{|\Omega_{t-1}|} \right)^{\frac{1}{1-\sigma}}}_{\text{Change in Variety Effects}} \times \underbrace{\left(\frac{|\Omega_t|/N_t}{|\Omega_{t-1}|/N_{t-1}} \right)^{\frac{1}{\theta}}}_{\text{Change in Selection Effects}}, \tag{A32}$$

where we can see that the first terms equal the growth from period $t-1$ to period t of the "Variety Effects" and "Selection Effects" terms derived in the ideal price index in the main paper. The "Average Price" term does not appear as the experiment of increasing N should leave unchanged the full price distribution and therefore there'd be no growth in $P^{\text{Conventional}}$.

The analysis is largely similar if we consider what one would obtain were they to mistakenly assume behavior in our environment was generated from a representative household and calculated a Feenstra correction using aggregate spending. As before, we'd have that $F_{1,t} = F_{3,t} = E$, as well as that $F_{2,t} = E \times N_{t-1}/N_t$. But now, we solve for $F_{4,t}$ by integrating the formula for the aggregate expenditure share – equation (22) – (times total expenditures) in period $t-1$ from the top good in terms of aggregate spending ($j = 0$) to the cutoff value of $\tilde{\gamma}_t^*$ which obtains at period t, which equals $j_t^* \times N_{t-1}/N_t$:

$$\begin{aligned}
F_{4,t} &= E \int_{j=0}^{j_t^* \frac{N_{t-1}}{N_t}} s_j dj = E \int_{j=0}^{j_t^* \frac{N_{t-1}}{N_t}} \frac{\eta+1}{\eta j_{t-1}^*} \left(1 - \left(\frac{j}{j_{t-1}^*} \right)^{\eta} \right) dj \\
&= E \frac{\eta+1}{\eta j_{t-1}^*} \left[j - \frac{1}{\eta+1} j^{\eta+1} \left(j_{t-1}^* \right)^{-\eta} \right]_0^{j_t^* \frac{N_{t-1}}{N_t}} \\
&= E \frac{\eta+1}{\eta} \left(\frac{j_t^*/N_t}{j_{t-1}^*/N_{t-1}} - \frac{1}{\eta+1} \left(\frac{j_t^*/N_t}{j_{t-1}^*/N_{t-1}} \right)^{\eta+1} \right) \\
&= E \frac{\eta+1}{\eta} \left(\left(\frac{|\Omega_t|/N_t}{|\Omega_{t-1}|/N_{t-1}} \right)^{\frac{1}{2}} - \frac{1}{\eta+1} \left(\frac{|\Omega_t|/N_t}{|\Omega_{t-1}|/N_{t-1}} \right)^{\frac{\eta+1}{2}} \right),
\end{aligned} \tag{A33}$$

where the last line substitutes in the relationship in equation (20). This then leads to the Feenstra correction one would make if they incorrectly assumed the data were generated by a single household:

$$F_t = \frac{\eta+1}{\eta} \frac{N_t}{N_{t-1}} \left(\left(\frac{|\Omega_t|/N_t}{|\Omega_{t-1}|/N_{t-1}} \right)^{\frac{1}{2}} - \frac{1}{\eta+1} \left(\frac{|\Omega_t|/N_t}{|\Omega_{t-1}|/N_{t-1}} \right)^{\frac{\eta+1}{2}} \right)^{-\frac{1}{\sigma-1}}. \tag{A34}$$

It is clear that, plugging the values for $|\Omega|$, N, and the other structural parameters into equations (A32) and (A34) will yield different answers. As discussed in the main text, equation (A32) is relevant for the household's welfare in our model, but equation (A34) is not since our model does not admit a representative agent with CES preferences.

B.2 Model Simulation Results

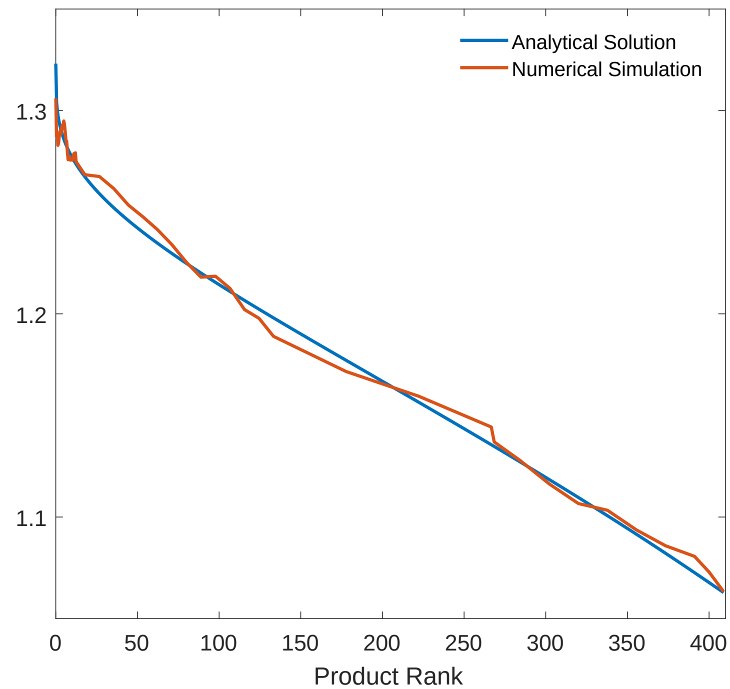
In this section, we explore numerical simulations of our model to test the validity of our elasticity approximation as well as to explore how restrictive the assumption of a stable distribution of Pareto taste-adjusted prices is for our conclusion.

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. Our 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 13. 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 16 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 $r_{i,j} = 0$, the second highest value to $r_{i,j} = 1$ and so on to $r_{i,j} = 2.25$ million. Finally, since for each $r_{i,j}$, 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 $r_{i,j} = 0$, 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 4.6, 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 A20 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 A20: 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.[13] 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 households 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 4.6.[14] 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.[15].

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 A21 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.[16] 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.

However, the welfare conclusion in the body of the paper under the assumed constant Pareto dis-

[13]Only 13.2% of UPCs purchased in 2004 are still purchased in 2016.

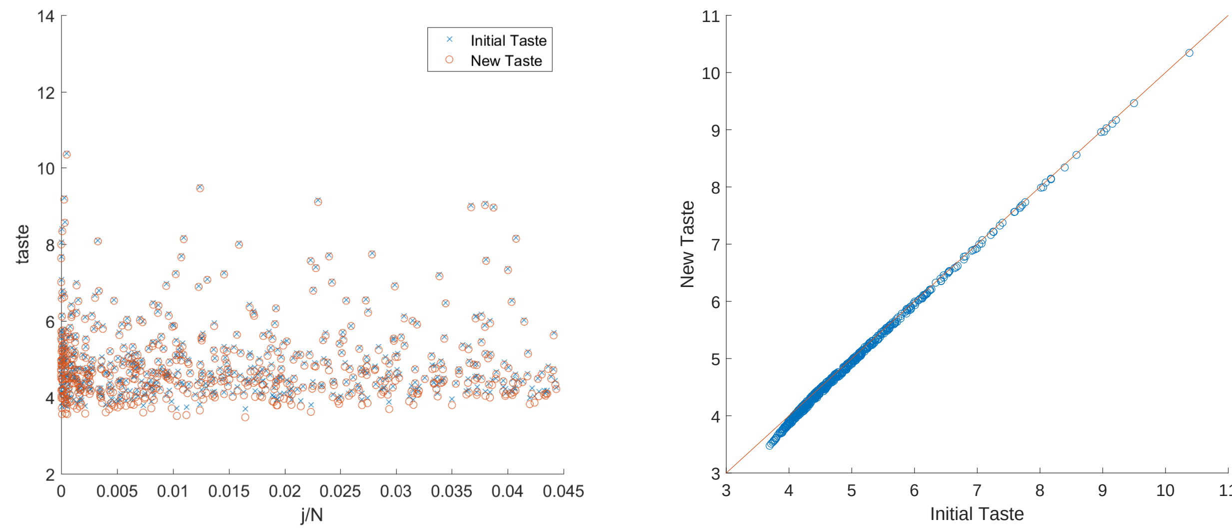
[14]For notational simplicity, we assume that marginal cost is 1 for all products. More generally this approach actually recovers the distribution of marginal cost adjusted tastes. As long as we assume marginal cost is constant as we vary N, one can interpret changes in taste and changes in marginal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

[15]Further, 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.

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

Figure A21: Household Taste Changes Required to Maintain Pareto with N increase

(a): Initial and New Tastes by Aggregate Product Rank **(b):** Initial vs. New Tastes



tribution of price-adjusted tastes is that an increase in N of 70% *raises* welfare by roughly 9.5%. The numerical results above show that 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 A.2 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. 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.

B.3 Specification with Linear Demand

In this subsection, we sketch a comparable setup to our baseline but using a linear demand system, following Melitz and Ottaviano (2008). We include key analytical results, such as derivations of Household and Aggregate Herfindhals, but additional details are available on request. We owe particular thanks to Levi Crews and Agustin Gutierrez for their outstanding research assistance in deriving the below expressions.

B.3.1 Setup

We assume household preferences are defined over a continuum of differentiated varieties indexed by $k \in (0, N]$. All consumers share the same utility function given by

$$U_i = \beta \int_{k\in(0,N]} \gamma_{i,k} C_{i,k} dk - \frac{1}{2}\sigma \int_{k\in(0,N]} (\gamma_{i,k} C_{i,k})^2 dk - \frac{1}{2}\eta \left(\int_{k\in(0,N]} \gamma_{i,k} C_{i,k} dk \right)^2,$$

where β, σ, η, and all preference shifters $\gamma_{i,k}$ are all non-negative. We impose an additional cost of consuming differentiated varieties, which we assume takes an exponential form in the measure of consumed varieties. Specifically, household i pays a cost $F \times (|\Omega_i|)^{\varepsilon}$ in units of the numeraire, where Ω_i is the set of differentiated varieties consumed in a positive amount by household i and F is non-negative. Household i therefore solves:

$$\max_{\{C_{i,k}\}} \; U_i \quad \text{s.t.} \quad \int_{k\in(0,N]} p_{i,k}C_{i,k}dk + F \times (|\Omega_i|)^{\varepsilon} \leq E.$$

We continue assume that price-adjusted tastes for each $k \in (0, N]$ are distributed Pareto for each household i with shape θ and support $[b, \infty)$ with $b > 0$. Because there is a continuum of varieties, each household faces the same set of taste-adjusted prices, though as before, their ranking of varieties within that set may differ. It can be shown that there exists a unique optimal measure of consumed varieties if and only if $\frac{\beta}{b}\left(\frac{\theta}{\theta+2}\right) N > E - F \times N^{\varepsilon}$. In the expressions that follow, assume this condition holds.

B.3.2 Key Expressions

Household expenditure shares in this environment can be written in closed-form as:

$$s_{i,k} = \frac{1}{|\Omega_i|}\left(\frac{(\theta+1)\,(\theta+2)}{\theta}\right)\left(1 - \frac{\tilde{\gamma}^*}{\tilde{\gamma}_{i,k}}\right)\frac{\tilde{\gamma}^*}{\tilde{\gamma}_{i,k}},$$

where $\tilde{\gamma}^* = |\Omega|^{-\frac{1}{\theta}} N^{\frac{1}{\theta}} b$. This then allows us to calculate and express the Household Herfindahl as:

$$\mathcal{H}^{HH} = \left[\frac{2(\theta+2)(\theta+1)^2}{\theta(\theta+3)(\theta+4)}\right]\frac{1}{|\Omega|}.$$

To move to the Aggregate Herfindahl, we impose the same form for the rank function as we used in the main analysis and obtain an expression for the aggregate share of product j:

$$s_j = \left[\frac{(\theta+1)(\theta+2)}{\theta}\right]\left[\frac{2(1-\alpha)}{\alpha N|\Omega|}\right]^{\frac{1}{2}}\left\{\frac{\theta}{\theta+2}\left[1-\left(\frac{j}{j^*}\right)^{\frac{2}{\theta}+1}\right] - \frac{\theta}{\theta+4}\left[1-\left(\frac{j}{j^*}\right)^{\frac{4}{\theta}+1}\right]\right\}.$$

This leads to the expression for the Aggregate Herfindahl:

$$\mathcal{H}^{\text{Agg}} = \frac{8}{3}\left[\frac{(\theta+1)(7\theta+12)}{(3\theta+4)(3\theta+8)}\right]\left[\frac{1}{2\tilde{N}|\Omega|}\right]^{\frac{1}{2}}.$$

B.3.3 Comparison to the CES Case

When we confront these expressions with the data to extract changes in θ and $\tilde{N}$, we find very similar results as what we found in the baseline CES case. In particular, the implied $\tilde{N}$ increases by roughly 70-80 percent, while the value for θ declines very slightly.

The implied value for θ, however, is much lower in the linear demand model than in the CES model. Though there is not a clear empirical benchmark for the value of θ, the value strikes as as less plausible than that from the CES case. Further, the aggregate market share distribution for such low values of θ do not conform as well with the CDF of market shares in our Nielsen data, as plotted in Figure 7. As a result, we use the CES specification as our benchmark.