# Looking into the Consumption Black Box: Evidence from Scanner Data*

Krzysztof Pytka[†] ⓡ Daniel Runge[‡]

February 2025

**Abstract**

This paper examines the heterogeneity and persistence of household non-durable consumption. We address three questions: (*i*) Do different consumer groups buy different products? (*ii*) How persistent are individual choices? (*iii*) What are the implications for structural models? We find minimal differences in basket composition between rich and poor households and high individual instability, with only 39% of products repurchased annually. To explain this, we propose a "shopping spree" model where products are perfect substitutes and baskets result from random sampling. Our findings serve as a cautionary note for structural models that emphasize product and consumer sorting.

**Keywords:** consumer choice, household heterogeneity, NielsenIQ, data sparsity

**JEL codes:** E00, E21, L11

---

*The order of authors is randomized, as indicated by the ⓡ symbol. We are thankful for comments made by Fernando Alvarez, Łukasz Drozd, Miklós Koren, Dirk Krueger, and Philipp Wangner. Lion Szlagowski's stellar research assistance was invaluable to this work. All errors are ours. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1597-1 FUGG. The researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[†]University of Mannheim; `me@kpytka.eu`

[‡]University of Mannheim; `drunge@mail.uni-mannheim.de`

In economics, there has been a notable shift away from the representative agent paradigm toward models that explicitly incorporate household heterogeneity. This trend reflects a growing recognition of the importance of accounting for individual differences in household decision-making. These differences are typically modeled as a result of intrinsic preferences that vary across consumers.

In this paper, we confront this underlying assumption on preferences using detailed non-durable consumption data from a large panel of households. To this end, we address three key questions: (*i*) Do different income groups of consumers purchase different products?; (*ii*) How persistent are individual choices over time?; (*iii*) Can those patterns be replicated by an alternative model of individuals where differences in consumption baskets are driven by different histories of product discovery rather than by systematic differences in preferences?

In answering these questions, the paper makes three main contributions. First, consumption choices are difficult to distinguish between rich and poor households, as spending patterns do not reveal a consumer's income level, suggesting that non-durable consumption is not polarized. Second, individual consumption choices are highly unstable, with only 39% of products purchased in one year being repurchased the following year. Finally, the paper proposes a parsimonious model of consumption, where products are treated as perfect substitutes, and basket composition results from random sampling. Remarkably, this model, which departs significantly from the standard approach in both its assumptions and implications, replicates observed consumption patterns surprisingly well. This finding serves as a cautionary note for models built on the assumption of hard-coded heterogeneity in consumption preferences.

All analyses in this article utilize the Kilts-NielsenIQ Consumer Panel (KNCP, henceforth). This dataset tracks 40,000–60,000 American households, capturing detailed scanner data. Our analysis spans from 2004 to 2016, covering 630 million transactions for around 800,000 barcode-level products annually across 87 million shopping trips.

To study consumption polarization we employ a two-stage strategy. First, we capture the differences between rich and poor households using an estimated multinomial model of consumption choices, where choices vary by income group. Subsequently, consumption polarization is measured by the average predictive power of the barcode of a product purchased with a randomly selected dollar within the KNCP universe. In highly polarized economies, where different income groups consume different products, the predictive power is expected to be high. Conversely, in less polarized economies, where different income groups consume similar products, the predictive power should be low.

Our identification strategy for polarization faces two methodological challenges. First, our dataset's choice space is overwhelmingly high-dimensional and our multinomial model, with around 800,000 categories, cannot be estimated using standard techniques. Second, despite the dataset's size, small-sample bias arises because the number of products far exceeds the number of consumers, making it difficult to distinguish genuinely polarizing goods from those consumed only once by chance.

To address these issues, we adapt the approach of Gentzkow, Shapiro, and Taddy (2019b), who faced a similar challenge in analyzing U.S. political polarization using congressional speech data. We mitigate small-sample bias by imposing a LASSO-type penalty on key income parameters and handle high dimensionality using the Poisson approximation of the multinomial model (Taddy, 2015), enabling distributed computing for feasible estimation.

We find that consumption polarization is much lower than commonly assumed. The way in which one dollar is spent allows us to predict whether a household belongs to the top or bottom decile of consumption expenditure with a probability of only 58.8%. Furthermore, this result remains stable, if not decreasing, over the studied time horizon. Aggregating into broader categories reduces this probability even further, approaching 50%, which represents the lower bound of the polarization measure.

Another dimension of consumption that we explore is the intertemporal stability of individual choices. To this end, we measure the persistence within a household's consumption bundle by computing the share of expenditures within a year spent on products already purchased in the previous year. We find that, on average, merely 39% of expenditures are spent on products that were purchased in the previous year, even after controlling for product entries and exits.

Our analysis suggests that the composition of consumption bundles is not influenced by income level; the choices made by high- and low-consumption households have almost no predictive power. This finding contrasts with models that rely on product sorting across the income distribution of consumers, suggesting that consumer choices *are not* driven by income. Additionally, the observed low stability of consumption baskets over time would require significant preference shocks in each period if modeled within a framework with latent heterogeneous preferences.

These insights motivate a thought experiment where we challenge the prevailing paradigm that differences in the composition of consumption baskets arise from heterogeneous preferences across households. To this end, we propose a deliberately *unconventional* modeling experiment that departs significantly from recent models based on intrinsic preference heterogeneity. Instead, our model of "shopping spree" assumes that all products are perfect substitutes (after adjusting for prices) and attributes variations in consumption baskets solely to random sampling. Our goal is to assess whether the empirical patterns reported in this article can be interpreted as the result of random sampling from a common distribution of products. It is important to emphasize that our objective is not to argue that this model provides a superior representation of reality. Rather, we seek to demonstrate that the observed patterns can be rationalized within a fundamentally different framework. This, in turn, serves as a cautionary tale for models that rely on intrinsic heterogeneous preferences. In this sense, the nature of our exercise is similar to that of Menzio (2024), who questions the Dixit-Stiglitz monopolistic competition model with a search-theoretic framework, or to Armenter and Koren (2014), who challenge the gravity model of international trade by introducing a model of random trade shipments.

Our model provides a thought-provoking experiment, demonstrating that a framework based on randomness and product substitutability can fit consumption data surprisingly well. Unlike models that emphasize heterogeneous preferences and product specialization by income groups, our approach challenges the necessity of such assumptions. While consumers may exhibit mild preferences for certain products, beyond the top-ranked choice, model predictions and observed data are nearly indistinguishable. This insight also raises questions about welfare analyses based on preference heterogeneity. For instance, the expansion of product variety at some cost would unambiguously reduce welfare in our model, which offers a different perspective from frameworks such as Neiman and Vavra (2023).

**Literature Review.** Our paper connects with several strands of economic literature. First, it contributes to the growing body of work that emphasizes the issue of data sparsity in large datasets. Gentzkow, Shapiro, and Taddy (2019b) highlight this problem in the text analysis of political speeches, while Armenter and Koren (2014) discuss the sparse nature of trade data and the surprisingly large class of trade models that are consistent with the available data.

Recently, economic models have increasingly focused on explicitly analyzing the consumer base, as seen in Afrouzi, Drenik, and Kim (2023) and Bornstein (2021). Our findings, which highlight the lack of consumption polarization and the low stability of individual baskets, provide direct modeling implications for this literature.

In the literature, the average price of a product within a class of products is often used as a proxy for quality, as discussed by Becker (2024) and Argente and Lee (2021). In our ongoing companion study (Runge ⓡ Pytka, 2025), we propose a simple yet powerful model experiment in which, within a Burdett and Judd (1983) search environment, all products are drawn from the same distribution, and high-income households have a lower probability of finding bargain deals than low-income households—consistent with findings in the empirical literature (Aguiar and Hurst, 2007; Kaplan and Menzio, 2015; Pytka, 2024).[1] Given the extent of data sparsity in the NielsenIQ universe, search frictions can generate a spurious quality ladder and consumption sorting across the income distribution, even though neither such ladder nor sorting exists in the original data-generating process.

[1] In our other study (Pytka, 2024), it is shown that data sparsity leads to an *underestimation* of price differentials across shoppers, while, in contrast, it may artificially amplify various consumption polarization measures. The latter is demonstrated using a permutation test on one measure of polarization: histogram overlap.

There is a vast literature studying heterogeneity in consumer preferences, exemplified by Handbury (2021), Neiman and Vavra (2023), and Michelacci, Paciello, and Pozzi (2021). This heterogeneity is modeled at varying levels of generality, with some models being more agnostic about its systematic sources, while others incorporate some additional factors such as search-and-discovery processes, as in Michelacci et al. (2021). Some recent structural models take a more explicit stance on consumption sorting across the income distribution, as seen in Nord (2023), Becker (2024), Sangani (2022), and Mongey and Waugh (2025). Our findings contribute to navigating these different theories and emphasize the importance of accounting for some randomness in consumer choices.

## I. Data Description

For this study, we use the Kilts-NielsenIQ Consumer Panel (KNCP) dataset to analyze price dynamics and consumption patterns. The KNCP tracks grocery purchases from a rotating panel of American households, expanding from approximately 40,000 households in 2004-2006 to around 60,000 from 2007 onward. Participants record purchases using in-home scanners or mobile apps, providing NielsenIQ with detailed transaction data from various retail outlets. Each purchase is linked to a specific shopping trip, and households submit socio-demographic information annually, with NielsenIQ assigning weights to ensure the sample reflects broader U.S. demographics. The dataset spans 54 geographic Scantrack markets and covers all available data from 2004 to 2016, including 630 million transactions involving nearly 2 million unique products (identified by UPCs) across 87 million shopping trips.

NielsenIQ classifies products into three levels of aggregation: department, group, and module. Each department consists of one or more groups, and each group contains one or more modules. For example, `FRUIT JUICE - GRAPEFRUIT - FROZEN` is a product module within the `JUICES, DRINKS-FROZEN` group, which belongs to the `FROZEN FOODS` department. We leverage these classifications to examine consumption patterns at different levels of granularity. Additionally, NielsenIQ provides a brand identifier linking individual products to broader brands, allowing us to group all products under a single brand where relevant.

## II. Empirical Patterns

In this section, we examine consumption behavior from two perspectives. First, we explore cross-sectional differences in the composition of consumption baskets between rich and poor households. Specifically, we ask whether knowing how a single dollar was spent — meaning which product it was used to purchase — provides a good predictor of a buyer's economic status. Second, we assess the stability of individual choices by examining whether past purchases increase the likelihood of repurchase.

### *A. Consumption Baskets of the Rich and Poor: Not So Different After All*

To study cross-sectional differences in the composition of consumption baskets between rich and poor households, we employ a polarization measure based on the predictive power of individual choices in determining group membership. In highly polarized societies, choices provide stronger predictive power than in less polarized ones.

In the recent literature, two prominent examples of prediction-based polarization measures are offered by Gentzkow, Shapiro, and Taddy (2019b) and by Bertrand and Kamenica (2023). In our analysis, we adopt the approach proposed by Gentzkow et al. (2019b), as we find it more suitable for our research question. However, where relevant, we will compare our findings to those of Bertrand and Kamenica (2023), particularly in sections where, like us, they examine consumption patterns.

The method proposed by Gentzkow et al. (2019b) consists of two steps. First, a model of consumption choices is estimated, allowing for differences across income groups. Second, the predicted choice distributions for the two groups from the first step are compared. More formally, let $\mathbf{c}_{i,t}$ be the observed $J$-dimensional consumption vector of individual $i$ at time $t$, which we assume comes from a multinomial distribution:

$$\mathbf{c}_{i,t} \sim \text{Multinomial}(m_{i,t}, \mathbf{q}_t^{P(i)}(\mathbf{x}_{i,t})) \tag{1}$$

where $\mathbf{c}_{i,t}$ represents the amount of money spent on each good by household $i$ in year $t$, and $m_{i,t}$ denotes the total expenditure of household $i$ in year $t$.[2] Here, $P(i)$ represents the income group to which household $i$ belongs (high $H$ or low $L$), $\mathbf{x}_{i,t}$ denotes a collection of household characteristics, and $\mathbf{q}_t^{P(i)}(\mathbf{x}_{i,t})$ refers to a set of choice probabilities with the following characteristics:

$$\begin{aligned} q_{jt}^{P(i)}(\mathbf{x}_{i,t}) &= \frac{\mathbf{e}^{\mathbf{u}_{i,j,t}}}{\sum_l \mathbf{e}^{\mathbf{u}_{i,l,t}}} \\ u_{i,j,t} &= \alpha_{j,t} + \mathbf{x}'_{i,t}\boldsymbol{\gamma_{j,t}} + \varphi_{j,t}\mathbf{1}_{i\in H_t} \end{aligned} \tag{2}$$

where $\alpha_{j,t}$ represents the baseline popularity of good $j$ in period $t$, $\gamma_{j,t}$ is a vector capturing the effect of household characteristics $\mathbf{x}_{i,t}$, and $\varphi_{j,t}$ captures the effect of belonging to the high-income group. This specification implies that the only household characteristics influencing choice probabilities are those included in $\mathbf{x}_{i,t}$, along with income group membership.

The vector of controls $\mathbf{x}_{i,t}$ includes household size, the age of the male and female household heads, and the presence and age of children.

The estimated model provides us with $\mathbf{q}_t^{P(i)}(\mathbf{x}_{i,t})$, representing the estimated distribution of choices for each household. From this perspective, all observed consumption choices in the NielsenIQ universe are realizations of the generative model given by Equation (2).[3]

Subsequently, given household characteristics $\mathbf{x}$, the difference in the composition of consumption baskets between rich $\mathbf{q}_t^H(\mathbf{x}_{i,t})$ and poor $\mathbf{q}_t^L(\mathbf{x}_{i,t})$ defines the

[2] In Appendix B, we replicate our analysis using quantities for $\mathbf{c}_{i,t}$ and $m_{i,t}$ instead of expenditures. The findings remain largely consistent with our baseline specification.

[3] A detailed discussion on the differences between generative and regression models can be found in Gentzkow et al. (2019a). While their analysis focuses on text data, the broader discussion applies to our context as well.

polarization measure.[4] When these vectors are similar, consumption baskets do not differ significantly across income groups. Gentzkow et al. (2019b) introduce a one-dimensional measure of polarization to capture the degree of divergence between multinomial distributions. In our application, it corresponds to the posterior probability that an observer with a neutral prior would assign to correctly identifying a shopper's income group based on the way a single dollar was spent:

$$\begin{aligned}\boldsymbol{\pi}_t(\mathbf{x}) &= \frac{1}{2}\mathbf{q}_t^H(\mathbf{x})\boldsymbol{\rho}_t(\mathbf{x}) + \frac{1}{2}\mathbf{q}_t^L(\mathbf{x})(1-\boldsymbol{\rho}_t(\mathbf{x}))\\ \rho_{i,t}(\mathbf{x}) &= \frac{q_{it}^H(\mathbf{x})}{q_{it}^H(\mathbf{x}) + q_{it}^L(\mathbf{x})}.\end{aligned}$$

Then, *average polarization* is given by:

$$\bar{\pi}_t = \frac{1}{|H_t \cup L_t|}\sum_{i\in|H_t\cup L_t|}\pi_t(x_{it}). \tag{3}$$

Average polarization measures the predictive power of knowing how a single dollar was spent in determining a shopper's income group. In a hypothetical scenario where rich and poor households consume exactly the same products, such as Coke, the predictive power would be 50%, meaning product information would be no more informative than a coin toss. Conversely, if rich and poor households consumed entirely different products—such as truffle-infused products and artisanal cheeses for the rich, and instant mac and cheese or spam for the poor—the predictive power would be 100%. In this case, product information would be as informative as knowing the true income group of the shopper. To illustrate how average polarization varies with different consumption patterns, we present a simulation in Appendix A.

For our application, a key advantage of the used measure is that it interprets polarization from the perspective of *aggregate consumption expenditures*. If certain products are consumed exclusively by rich or poor households but account for a negligible share of total spending, the average polarization measure remains low.

[4] In the original paper, the authors use the term *partisanship* due to its political context. Here, we adopt the term *polarization* as it provides a more general interpretation.

This perspective complements the approach of Bertrand and Kamenica (2023), who examine whether there *exists* a set of products that can predict a shopper's income group, irrespective of its contribution to total spending. While their method provides insights into the existence of predictive product sets, our approach instead summarizes the *information value of a randomly selected single dollar spent*, capturing how income-related differences in consumption expenditures manifest at the aggregate level.

Before presenting the results, we first highlight the key challenges in our analysis. The first challenge is the high dimensionality of the choice set. At the most granular product definition—the barcode level—the number of categories is approximately 800,000 every year. This makes the estimation of the multinomial model in Equation (2) numerically infeasible using standard econometric techniques. We address this issue by employing a Poisson approximation to the multinomial model, as proposed by Taddy (2015). This approximation enables distributed computing, making the estimation procedure computationally feasible.

The second challenge is data sparsity. The NielsenIQ panel data is highly sparse, which can lead to severe small-sample bias and spurious polarization. In our application, the number of product categories is significantly larger than the number of panelists for most product definitions. As a result, many products are consumed by only a small number of households, making it difficult to determine whether they are genuinely polarizing goods or simply consumed once by chance.

To illustrate the magnitude of this issue, as we document it in another study (Pytka, 2024), 30% of aggregate monthly consumption consists of transactions involving products purchased by only one household. Many of these products, which may be bought only once by chance, would act as perfect predictors—not only of income group but of all household characteristics, even down to an individual's social security number. Consequently, the estimated model would suffer from severe overfitting.[5] To mitigate this problem, we regularize $\varphi_{j,t}$ in Equa-

[5]We also examine transaction frequency on an annual basis (Runge ⓡ Pytka, 2025) and find that 50% of aggregate consumption expenditures come from products purchased fewer than 200 times per year. This suggests that small-sample concerns remain highly relevant even with annual aggregation.

tion (2), which accounts for income-group membership, using a LASSO penalty. The optimal penalty value is selected based on an information-based criterion. This approach allows us to identify genuinely polarizing products while filtering out those consumed purely by chance.

In our baseline analysis, we define rich and poor households as those in the top and bottom quintiles of consumption expenditure, respectively. As a robustness check, detailed in Appendix B, we replicate our analysis using income-bracket information instead.[6]

Additionally, we consider several definitions of goods. At increasing levels of granularity, we define a product at the department, group, module, and barcode (UPC) levels. In this section, we primarily focus on results based on the barcode definition while briefly discussing findings for the other definitions. A more detailed analysis, including graphical representations, is provided in Appendix B. Due to computational constraints, we do not report confidence bands for the polarization measure at the barcode level.[7] For the same reason, we conducted the polarization analysis at the most granular level—the barcode—by first examining each module separately and then aggregating the results to obtain the average barcode-level polarization. The exact procedure is described in Appendix C, while barcode-level polarization within each department is presented in Appendix D.

Figure 1 illustrates the dynamics of consumption polarization over time for products defined at the barcode level. The average polarization measures for the years 2004–2016 can be summarized as follows:

**Fact 1. (Consumption Polarization)** *The average consumption polarization* $\bar{\pi}$ *is low and close to its lower bound of 50%. Specifically, if we randomly draw \$1 spent by a high- or low-income household in the NielsenIQ universe and observe*

[6]However, NielsenIQ provides income-bracket data with a two-year delay. We address this issue by matching income information with consumption data from different waves of the dataset. Nonetheless, this results in a smaller panel sample, as some panelists exit the dataset. For this reason, we rely on the consumption-based definition of rich and poor households in the main text.

[7]Although confidence bands are unavailable at the lowest level of granularity, our estimates remain stable over time, suggesting that they are reasonably accurate. Moreover, polarization measures at higher levels of granularity are estimated with such small confidence bands that they are not visible in the figures.

*how it was allocated, the probability of correctly classifying the spender is:*

*i. 58.3% when the **barcode** of the purchased product is observed,*

*ii. 52.2% when the **product module** is observed,*

*iii. 51.7% when the **product group** is observed,*

*iv. 50.6% when the **product department** is observed.*

These results indicate that consumption sorting between rich and poor households is weak, as product choices across different income groups exhibit substantial overlap, even at detailed classification levels.

We observe several common patterns across all estimated specifications. First, polarization remains relatively stable over time, showing neither an upward nor a downward trend. By construction, polarization decreases as the level of aggregation increases.[8] At the product department level, polarization is very low, with a maximum value of 51%. Even at the product group or module level, polarization remains minimal, suggesting that distinguishing between income groups based solely on consumption choices is nearly impossible at these levels.

The fact that we observe very low levels of polarization even at the most granular level might appear to stand in contrast to Bertrand and Kamenica (2023), who find that consumption choices allow for forecasting a household's income group with approximately 90% accuracy. Beyond differences in data sources, this discrepancy can be attributed to methodological differences in how polarization is defined. We view our approaches as complementary rather than conflicting. While the study by Bertrand and Kamenica (2023) is conducted from the perspective of cultural economics—where even small differences in consumption, such as the *existence* of products consumed exclusively by one group, may have a meaningful impact on individual attitudes and beliefs—our approach focuses on aggregate consumption patterns, specifically the amount of information contained in a "representative" dollar spent, and market-based interactions between economic entities.[9]

[8] At the extreme, if all products were considered a single good, the average polarization would be 50%.

[9] Later in Subsection C of the current section, we highlight two examples, price discrimination

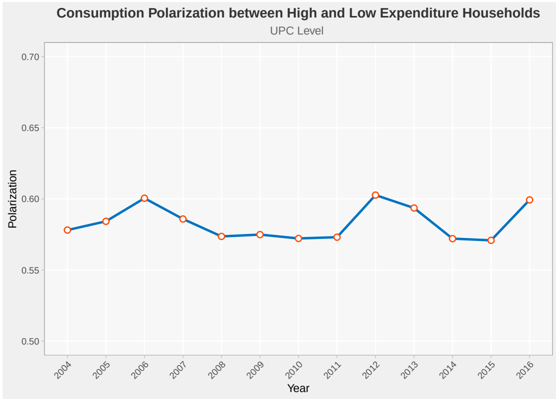



Figure 1: The plot displays polarization estimates between the top 20% and bottom 20% of the consumption expenditures distribution. The estimates are based on the within-department estimates. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

*B. Individual Consumption Baskets: Constantly in Flux*

We examine the dynamic stability of consumption by calculating the probability that a product in a household's basket remains there over time. This measure of persistence provides insights into customer-brand relationships at the barcode or brand level and into households' preference stability at the module level. Our primary measure of consumption persistence is the expenditure-weighted average probability that a product remains in the basket over time. More precisely, let $\mathcal{U}_{i,t}$ denote the set of products purchased by household $i$ in year $t$, and let $e_{i,t}(j)$ represent household $i$'s expenditure on product $j$ in year $t$. The expenditure-weighted persistence, $O^E_{i,t+1}$, is then defined as:

$$O^E_{i,t+1} = \frac{\sum_{j \in (\mathcal{U}_{i,t} \cap \mathcal{U}_{i,t+1})} e_{i,t+1}(j)}{\sum_{j \in \mathcal{U}_{i,t+1}} e_{i,t+1}(j)}. \tag{4}$$

This formulation captures the share of expenditures in year $t+1$ allocated to products that were also purchased in year $t$, providing a measure of persistence in

and search externalities, where the polarization measures reported by Bertrand and Kamenica (2023) may not fully capture the relevant aspects of polarization for these cases.

household consumption patterns. We consider three levels of product definition for $j$ : barcode level, brand-module, and module level.

Figure 2 illustrates the dynamics of persistence levels. The average persistence measures for years 2004-2016 can be summarized as follows:

**Fact 2. (Persistence in Expenditures)** *The average persistence in household expenditures varies by product definitions. Specifically, the share of expenditures on products purchased in the previous year is, on average:*

*i. 38.8% for products defined at the* ***barcode*** *level,*

*ii. 59.5% for products defined at the* ***brand*** $\times$ ***module*** *level,*

*iii. 83.5% for products defined at the* ***module*** *level.*

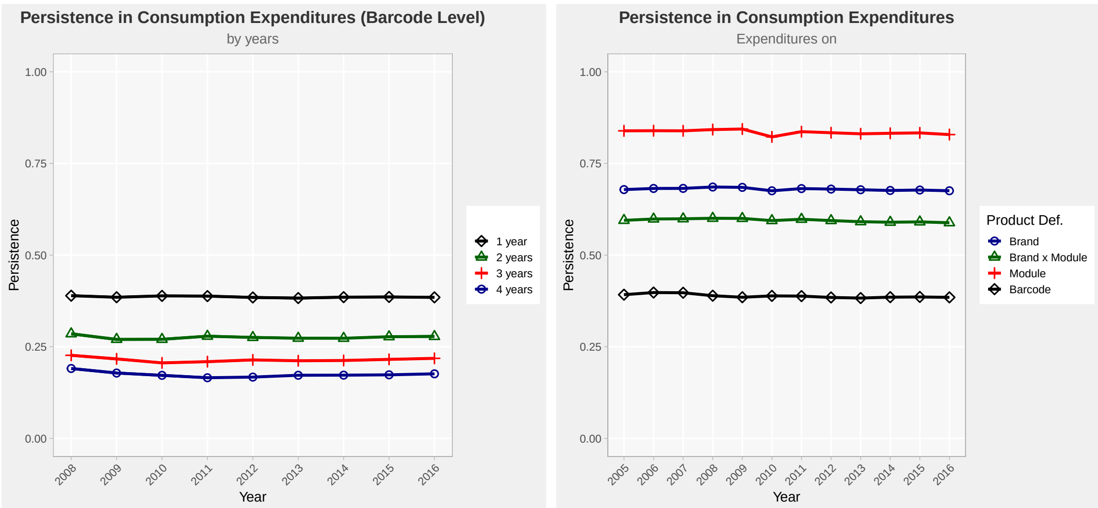



Figure 2: The plot on the left side shows barcode-level persistence estimates for different time horizons and the plot on the right persistence estimates for different product definitions. All confidence bands are 95% and are not visible because they are very narrow compared to the scale of the y-axis which is kept identical for all plots in this section.

There are two key takeaways from our finding. First, the average persistence level is relatively high at the module level. The share of expenditures allocated to specific product types (where product module granularity defines goods such as peanut butter, herbal tea in bags, or natural American cheddar) remains stable over time. On the other hand, persistence at the barcode level is strikingly low. This suggests that while households consistently allocate similar proportions of their expenditures to the same product categories, they frequently switch between specific products.[10]

One potential explanation for the low persistence at the barcode level is product entry and exit. However, this is not the case. In Appendix E, we repeat the analysis, excluding products that exit the market, and find results consistent with our main findings.

An alternative perspective on the high instability of consumption baskets at the barcode level comes from search theory, which suggests that products remaining in the basket are those found at retailers offering lower prices, leading to greater stability in customer-product relationships. However, our findings contradict this interpretation. On average, goods that remain in the basket are purchased at prices *0.18% above* their mean, while those that are dropped are bought at prices *0.07% below* their average. Not only are these differences minimal, but their direction is also opposite to what price-search models would predict. This suggests that price alone does not explain the instability of consumption baskets as expected.

To investigate whether products leave baskets temporarily or permanently, we recomputed the barcode persistence measure using 2-, 3-, and 4-year gaps. Figure 2 shows that persistence decreases with longer delays, suggesting that products removed from the basket are unlikely to return in the future.

Our result on the low stability of consumption baskets might seem in stark contrast to the recent findings of Bornstein (2021), who reports that shoppers leave a *firm*'s consumer base with an average probability of around 16%. However, our results are not directly comparable to his, as we focus on more granular definitions of products and baskets, whereas he examines the multi-product firm level.

Next, in our analysis we examine heterogeneity in individual (barcode level) persistence across households. As Figure 3 illustrates, the variation is striking. Moreover, no socio-demographic characteristics are correlated with individual per-

[10] As a robustness check, we conducted a similar analysis in which expenditures in Equation 4 were replaced with products. Qualitatively, the findings remain the same, with even lower basket stability. The results are presented in Appendix E.

sistence. We trained a random fores to predict persistence using a set of socio-demographic characteristics provided by NielsenIQ, but found no meaningful correlations. This suggests that individual persistence is not driven by observable characteristics. Since our analysis reveals no significant correlations, we have relegated the detailed results to Appendix F.

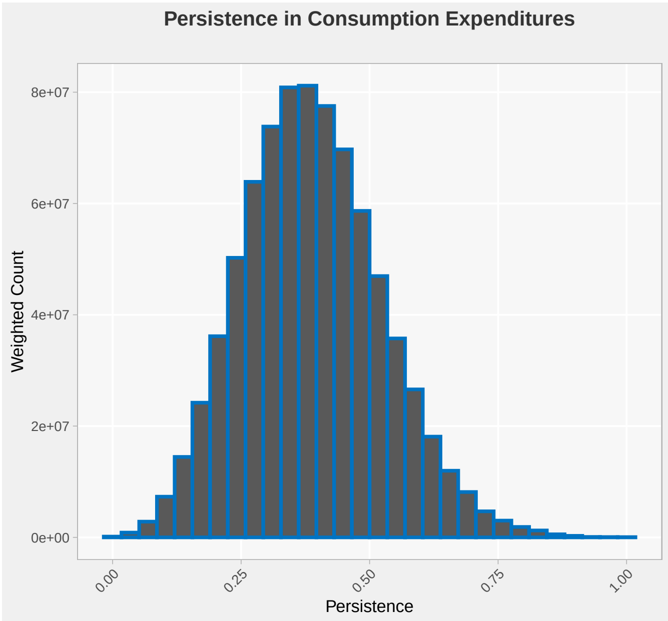



Figure 3: Histograms of persistence estimates. The figure on the left shows the histogram of persistence estimates using the UPC based measure and the one on the right the expenditure based one. In both figures the estimates are weighted using the projection factors.

### *C. Taking Stock*

Our empirical analysis reveals key insights into household consumption behavior. We find that consumption polarization is low, as product choices across income groups exhibit substantial overlap. Additionally, individual consumption choices appear highly unstable over time, with only 38.8% of products being repurchased annually. While households frequently change the specific items they buy at the barcode level, their broader consumption composition remains more stable at the product module level. These findings highlight the fluidity of household shopping behavior and suggest that persistent, well-defined preferences might play a less important role in shaping long-term consumption patterns than commonly assumed.

The low level of consumption sorting, as documented in **Fact 1**, has several immediate implications for modeling consumption choices. Here, we highlight two illustrative examples. First, price discrimination — where producers offer different products to different income groups at prices tailored to highly specific consumer segments — appears challenging in light of our results. Second, due to the significant overlap in purchased goods, substantial price search externalities are very plausible, where the search behavior of one group affects shopping constraints and decision-making of others, as described in the consumer search model by Pytka (2024).

The empirical evidence on the stability of individual consumption baskets, as documented in **Fact 2**, reveals a nuanced pattern of consumer behavior. Households exhibit strong preferences for certain types of products, as reflected in the high stability at the product module level, where 83.5% of expenditures are allocated to previously purchased categories. However, at the barcode level, this stability drops to just 38.8%, indicating that attachment to specific products is limited. While some brand loyalty exists within product modules, it remains below 60%, suggesting that consumers frequently switch between brands rather than consistently repurchasing the same items. Moreover, the absence of systematic price differences between products entering and leaving the basket further supports the idea that consumer-firm relationships at the barcode level are ephemeral, with households regularly adjusting their exact product choices while maintaining broader category preferences. One immediate implication of this result for modeling is that firms' expansion strategies should be viewed primarily as the acquisition of new customers rather than the retention of existing ones, as in the framework proposed by Afrouzi, Drenik, and Kim (2023).

## III. A Model of Shopping Spree

In this section, we propose a model that challenges the notion that differences in consumption baskets stem from heterogeneous preferences. Instead, our framework assumes all products are perfect substitutes (after adjusting for prices), with basket variations arising purely from random sampling. Our goal is not to present a more realistic model but to demonstrate that the observed patterns can emerge in a fundamentally different setting.[11]

In our parsimonious model, households make purchasing decisions during a *"shopping spree."* Unlike models in which households have intrinsic preferences for specific goods and select those they most prefer, we assume that all products are perfect substitutes (adjusted for prices).[12] Moreover, motivated by **Fact 1**, we assume no product sorting toward specific consumer groups, meaning that while consumption probabilities vary across products, each product is consumed with the same probability across all households.

Formally, let $i \in I$ denote a household in the NielsenIQ universe, $I$. Each consumer is characterized by their annual consumption spending in year $t$, denoted by $m_{it}$, with no possibility of saving. Given the budget constraint and assumed preferences, each household maximizes its total expenditures. The actual composition of products in their baskets is irrelevant to their utility. Instead, each household samples the composition of its basket. While all households draw from the same marginal distribution of products, the probability of selecting a product (defined at the barcode level) $j$ and the quantity of product $j$ in the basket are determined by a product-specific zero-inflated Poisson distribution.[13] This means that random

[11]In this sense, our model aim at serving as a cautionary tale for models relying on intrinsic preference heterogeneity, much like Menzio (2024) and Armenter and Koren (2014), who challenge other popular frameworks (monopolistic competition and gravity models of international trade, respectively) with alternative mechanisms.

[12]This means that, on average, more expensive products provide higher utility, but households are indifferent between purchasing one unit of a product that is twice as expensive as buying two units of a cheaper alternative. Price serves as a perfect summary of utility.

[13]We remain agnostic about how these probabilities are determined—they could arise from prices, product comparisons, or marketing influences. The key assumption is that these probabilities are *equal* across different income groups, ensuring that no systematic sorting of products occurs based on economic status.

sampling determines whether product $j$ is included in the basket of household $i$ in period $t$, and if it is included, how many units of product $j$ are purchased. Overall, each household's consumption choices are summarized by a $J$-dimensional vector $\mathbf{c}_{i,t}$, where entry $j$ represents the number of units consumed of good $j$.

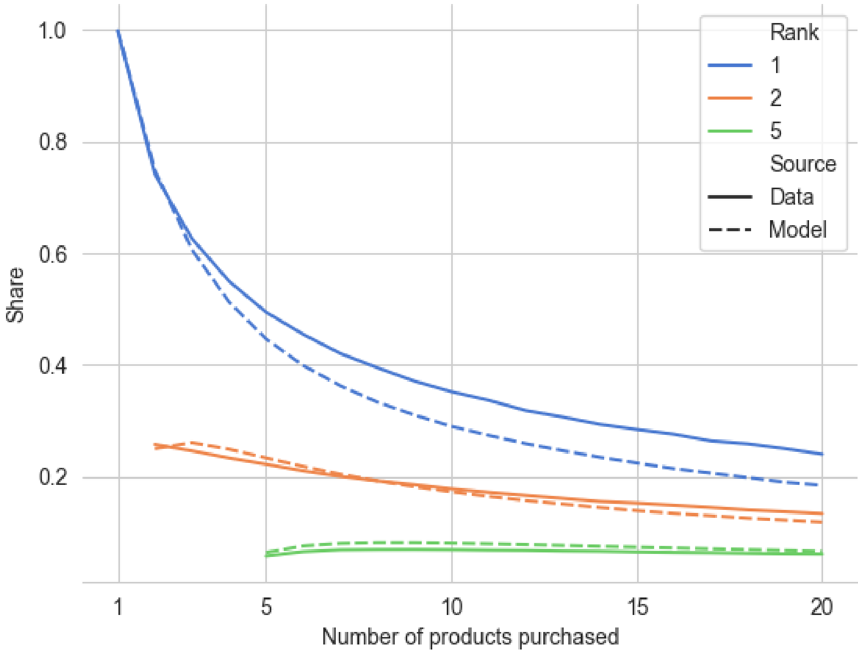
In our simulation, we assume that the unit price of each product equals the average transactional price for that product in the NielsenIQ universe. A simulation of the shopping process for household $i$ is summarized in Algorithm 1.[14]

**Algorithm 1** Shopping Spree Model

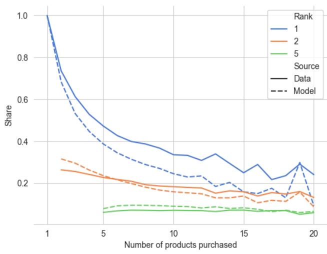
1. **Initialize:** Set consumption vector $\mathbf{c}_{i,t} = \mathbf{0}$.
2. **while** budget constraint is not violated ($\mathbf{p}'_{i,t}\mathbf{c}_{i,t} < m_{i,t}$) **do**
3. Randomly draw a product $j$ from the set of all products, with a *product-specific* probability that is the same across all consumers.
4. Draw the number of units purchased, $n_j$, from a *product-specific* Poisson distribution. Exclude product $j$ in future draws of households $i$.
5. Update consumption vector: $\mathbf{c}_{i,t} \leftarrow \mathbf{c}_{i,t} + \mathbf{e}_j n_j$, where $\mathbf{e}_j$ is a unit vector with 1 at the $j$-th position and 0 elsewhere.
6. **end while**
7. **Stop.**

The model outlined here can be seen as a data-driven, high-dimensional extension of the classical model of impulsive customers proposed by Becker (1962). Compared to that model, we expand the consideration set to include all barcode-level products in the NielsenIQ universe, rather than just two as in the original study, and estimate probabilities directly from the data. On the other hand, our implementation closely resembles the "balls-and-bins" model of international trade by Armenter and Koren (2014), which introduces a simple, atheoretical random-assignment approach based solely on marginal distributions across categories (in their case, trade distribution across countries or products), without requiring information on specific country-product trade links or assuming systematic trade relationships. Similarly, in our model, consumer baskets emerge as the result of random assignment.

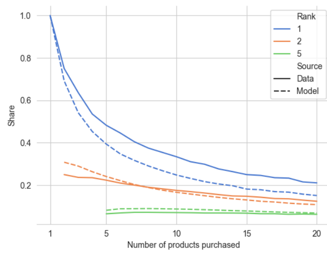
[14] Admittedly, it is possible for households to slightly violate their budget constraint, $m_{it}$. However, given the scale of annual expenditures and the spending on individual products, the magnitude of these violations is negligible.



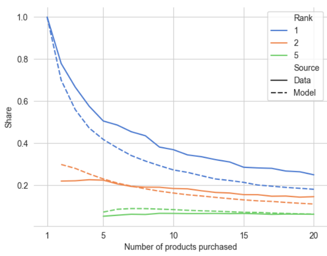

(a) Average Spendings on Different Ranked Items



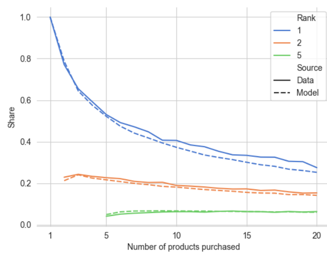

(b) Seafood Canned




(c) Cereal




(d) Yogurt




(e) Pet Food

Figure 4: Comparison of Average Spendings and Category Spendings on Different Ranked Items

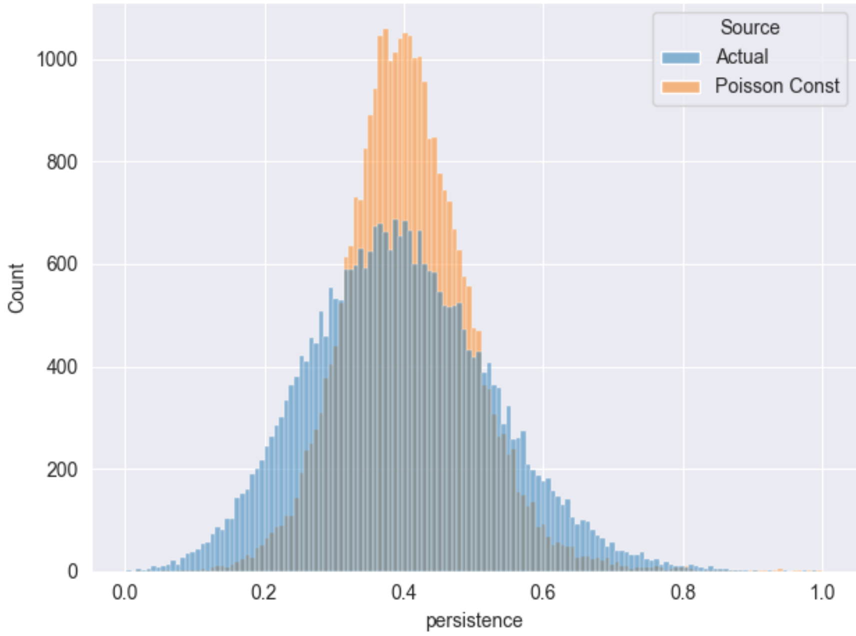



Figure 5: Household Consumption Persistence

Figure 4a illustrates the relationship between household spending shares on their top-, second-, and fifth-ranked goods and the total number of goods consumed. The solid lines represent the empirical averages, weighted by total spending, while the dashed lines show the theoretical predictions of the shopping spree model. The construction of these plots follows the methodology of Neiman and Vavra (2023), for whom this plot constitutes the main validation moment at the household level. When households purchase only one good, it naturally accounts for 100% of their spending, aligning both model and data. As consumption diversifies, the top good's share declines, reaching around 20% for households purchasing 20 products. Similarly, spending shares on the second- and fifth-ranked goods decrease, following patterns closely mirrored by the model. Figures from 4b to 4e repeat this analysis within product departments, using the same categorization as in Neiman and Vavra (2023).

Despite its simplicity, our proposed model captures the main features of consumer behavior data remarkably well. However, unlike models based on heterogeneous preferences, it relies on randomness and treats consumers as impulsive agents, considering all products as perfect substitutes. This approach contrasts starkly with traditional consumer theory while still yielding empirically consistent results. Admittedly, there is a small discrepancy between the share of the top-ranked product in the model and the data, suggesting that consumers do have some preferences for their favorite products. However, beyond the top-ranked product, the differences between model predictions and observed data become practically indistinguishable, starting from the second-most preferred product onward.

Our model serves as a thought-provoking experiment, demonstrating that a framework fundamentally different from standard heterogeneous-preference models fits the data surprisingly well. This finding challenges theories where product specialization by income groups is central. Likewise, the absence of systematic preferences for specific products challenges models in which the welfare effects of policies are primarily driven by preference heterogeneity. For instance, in our framework, introducing costs to expand variety would unambiguously reduce welfare, whereas models that emphasize heterogeneous preferences, such as Neiman and Vavra (2023), suggest that such policies could have welfare-enhancing effects.

While our model captures cross-sectional patterns relatively well, persistence in a dynamic setup—without additional components—would be even lower than the lowest observed value at the barcode level. To address this, we introduce an ad-hoc persistence parameter, $\rho = 38.8\%$, representing the probability that a purchased product will be repurchased in the next period. This extension can be interpreted as "inertia," similar to the assumption made by Becker (1962). All previously reported cross-sectional characteristics remain unchanged. Unsurprisingly, this extension increases consumption persistence, bringing the first moment precisely to the calibrated 38.8%. More strikingly, the simulation also generates a level of heterogeneity in consumption persistence that closely matches the empirical distribution seen in Figure 5. In the simulation, dispersion in persistence arises from differences in the number of transactions—households with more transactions exhibit much lower variation in persistence. Given the lack of correlation between persistence and observable characteristics, along with the similar pattern emerging in our simulation, this heterogeneity in persistence may be a statistical artifact driven by variation in the number of draws. In Appendix E, Figure 20 confirms this, showing that households with fewer transactions exhibit greater dispersion in persistence.

## IV. Concluding Thoughts

Our analysis of non-durable consumption behavior has provided several insights into household decision-making. We find that consumption patterns across income groups show minimal polarization, with substantial overlap in the products purchased by rich and poor households. This suggests that, contrary to models emphasizing consumption sorting, the composition of consumption baskets is more homogeneous than often assumed.

Furthermore, the high instability of individual consumption baskets—only 39% of products are repurchased annually—underscores the transient nature of choices. This instability challenges the idea of stable systematic heterogeneity in preferences, suggesting that random variation plays an important role in consumption decisions.

While our results challenge the standard view, we recognize that some randomness in consumer choices has already been incorporated, such as in Michelacci et al. (2021), where heterogeneous preferences are combined with search-and-discovery processes.

Further critical exploration of the concept of a quality ladder, particularly in light of data sparsity, warrants additional research. In this context, our ongoing companion study (Runge ⓡ Pytka, 2025) proposes a simple yet powerful model experiment in which search frictions, combined with data sparsity, can generate a spurious quality ladder.

## References

Afrouzi, Hassan, Andres Drenik, and Ryan Kim (2023) "Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition," May.

Aguiar, Mark and Erik Hurst (2007) "Life-Cycle Prices and Production," *American Economic Review*, 97 (5), 1533–1559, http://ideas.repec.org/a/aea/aecrev/v97y2007i5p1533-1559.html.

Argente, David and Munseob Lee (2021) "Cost of living inequality during the great recession," *Journal of the European Economic Association*, 19 (2), 913–952.

Armenter, Roc and Miklós Koren (2014) "A Balls-and-Bins Model of Trade," *American Economic Review*, 104 (7), 2127–2151, 10.1257/aer.104.7.2127.

Becker, Gary S. (1962) "Irrational Behavior and Economic Theory," *Journal of Political Economy*, 70 (1), 1–13.

Becker, Jonathan (2024) "Do Poor Households Pay Higher Markups in Recessions?", Working paper.

Bertrand, Marianne and Emir Kamenica (2023) "Coming Apart? Cultural Distances in the United States over Time," *American Economic Journal: Applied Economics*, 15 (4), 100–141, 10.1257/app.20210663.

Bornstein, Gideon (2021) "Entry and profits in an aging economy: The role of consumer inertia,"Technical report.

Burdett, Kenneth and Kenneth L Judd (1983) "Equilibrium Price Dispersion," *Econometrica*, 51 (4), 955–69, http://ideas.repec.org/a/ecm/emetrp/v51y1983i4p955-69.html.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019a) "Text as Data," *Journal of Economic Literature*, 57 (3), 535–574, 10.1257/jel.20181020.

_ , Jesse M. Shapiro, and Matt Taddy (2019b) "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 87 (4), 1307–1340, https://doi.org/10.3982/ECTA16566.

Handbury, Jessie (2021) "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," *Econometrica*, 89 (6), 2679–2715, https://doi.org/10.3982/ECTA11738.

Kaplan, Greg and Guido Menzio (2015) "The Morphology of Price Dispersion," *International Economic Review*.

Menzio, Guido (2024) "Markups: A Search-Theoretic Perspective," NBER Working Paper 32888, National Bureau of Economic Research, Available at `https://www.nber.org/papers/w32888`.

Michelacci, Claudio, Luigi Paciello, and Andrea Pozzi (2021) "The Extensive Margin of Aggregate Consumption Demand," *The Review of Economic Studies*, 10.1093/restud/rdab041.

Mongey, Simon and Michael E. Waugh (2025) "Pricing Inequality," NBER Working Paper w33399, National Bureau of Economic Research, Available at `https://www.nber.org/papers/w33399`.

Neiman, Brent and Joseph Vavra (2023) "The Rise of Niche Consumption," *American Economic Journal: Macroeconomics*, 15 (3), 224–264.

Nord, Lukas (2023) "Shopping, Demand Composition, and Equilibrium Prices,"Technical report.

Pytka, Krzysztof (2024) "Shopping Frictions with Household Heterogeneity: Theory  Empirics,"Technical report.

Runge, Daniel ⓡ Krzysztof Pytka (2025) "The Mirage of Consumption Sorting," Working paper.

Sangani, Kunal (2022) "Markups across the income distribution: Measurement and implications," *Available at SSRN 4092068*.

Taddy, Matt (2015) "Distributed Multinomial Regression," *The Annals of Applied Statistics*, 9 (3), 1394–1414, 10.1214/15-AOAS848.

— and Nelson Rayl (2022) "Package 'distrom'," https://github.com/TaddyLab/distrom, Version 1.0.1.

# A. Online Appendix

*A. Simulation*

We provide a simple simulation exercise to highlight how our polarization measure works. The goal is to establish a benchmark for estimating polarization and interpreting the results derived from it. We simulate an economy with 10 different goods. It is populated with 80,000 households, each consuming 200 units of goods in each period. The 200 units consumed by each household are drawn randomly given a fixed set of probabilities. The economy is simulated for one period, and the true choice probabilities for the various simulations are as follows:

| **Preference Type** | **Probabilities** | **Good 1** | **Good 2** | **Good 3** | **Good 4** |
|---|---|---|---|---|---|
| Homogeneous Preferences | Uniform Probability | 0.1 0.1 | 0.1 0.1 | 0.1 0.1 | 0.1 0.1 |
| | Non-Uniform Probability | $\frac{0}{18}$ $\frac{0}{18}$ | $\frac{1}{18}$ $\frac{1}{18}$ | $\frac{2}{18}$ $\frac{2}{18}$ | $\frac{3}{18}$ $\frac{3}{18}$ |
| Heterogeneous Preferences | Perfect Separation | 0.2 0.0 | 0.2 0.0 | 0.2 0.0 | 0.2 0.0 |
| | Partial Separation | $\frac{0}{18}$ $\frac{9}{18}$ | $\frac{1}{18}$ $\frac{8}{18}$ | $\frac{2}{18}$ $\frac{7}{18}$ | $\frac{3}{18}$ $\frac{6}{18}$ |

| **Preference Type** | **Probabilities** | **Good 5** | **Good 6** | **Good 7** | **Good 8** |
|---|---|---|---|---|---|
| Homogeneous Preferences | Uniform Probability | 0.1 0.1 | 0.1 0.1 | 0.1 0.1 | 0.1 0.1 |
| | Non-Uniform Probability | $\frac{4}{18}$ $\frac{4}{18}$ | $\frac{5}{18}$ $\frac{5}{18}$ | $\frac{6}{18}$ $\frac{6}{18}$ | $\frac{7}{18}$ $\frac{7}{18}$ |
| Heterogeneous Preferences | Perfect Separation | 0.2 0.0 | 0.0 0.2 | 0.0 0.2 | 0.0 0.2 |
| | Partial Separation | $\frac{4}{18}$ $\frac{5}{18}$ | $\frac{5}{18}$ $\frac{4}{18}$ | $\frac{6}{18}$ $\frac{3}{18}$ | $\frac{7}{18}$ $\frac{2}{18}$ |

| **Preference Type** | **Probabilities** | **Good 9** | **Good 10** |
|---|---|---|---|
| Homogeneous Preferences | Uniform Probability | 0.1 0.1 | 0.1 0.1 |
| | Non-Uniform Probability | $\frac{8}{18}$ $\frac{8}{18}$ | $\frac{9}{18}$ $\frac{9}{18}$ |
| Heterogeneous Preferences | Perfect Separation | 0.0 0.2 | 0.0 0.2 |
| | Partial Separation | $\frac{8}{18}$ $\frac{1}{18}$ | $\frac{9}{18}$ $\frac{0}{18}$ |

Table 1: Choice probabilities for both types of households. The probabilities for the first type are in red and for the second in cyan.

The four different sets of choice probabilities for the simulations represent different illustrative scenarios for selection patterns between the two groups. The first two scenarios represent cases where preferences are homogeneous. In scenario one, the probabilities are uniform across both groups and goods, while in scenario two, they are non-uniform. Scenarios three and four represent cases of heterogeneous preferences. In scenario three, the separation between the two groups is perfect, while in scenario four, the separation is imperfect.

In the first scenario, preferences are homogeneous and both groups select any of the ten goods with the same probability. This leads to consumption baskets that are, on average, identical, with each good having the same share in the basket. Therefore, households are indistinguishable based solely on their purchases, which should yield a polarization estimate of 0.5 since the group prediction is equivalent to a coin flip.

In the second scenario, preferences are again homogeneous, but choice probabilities are not identical among goods. Both groups are least likely to buy good 1, with the probability increasing and being highest for good 10. The average consumption baskets for the two groups will again be identical in this case, with each good having a different share in the basket. Similar to Scenario 1, households cannot be distinguished based solely on their choices, since the product shares within the baskets are, on average, identical between the two groups. Therefore, the polarization measure should be equal to 0.5 in this case.

In the third scenario, preferences are heterogeneous. The first group will only purchase the first five goods, while the second group will purchase the remaining five goods, each with equal probability. Therefore, the average consumption baskets for the two groups will share no common goods, while each good included in a basket will have the same share. Since there is no overlap in consumption baskets between the two groups, households are perfectly distinguishable based on consumption choices, and therefore the polarization measure will be 1.

In the final scenario, the choice probabilities for the first group are the same as in the third scenario. For the second group, the probabilities are exactly reversed, meaning they are least likely to purchase good 10 and most likely to purchase good 1. In this case, the average baskets for both groups will contain the same goods, but the shares will be different for the two groups. Thus, a choice for one of the products is informative about which group the household belongs to, but group membership cannot be perfectly deduced from it. Notably, even though goods 1 and 10 are perfect predictors of group membership, polarization will not be equal to 1. This is because our measure captures the average information contained in a purchase. In this example, the most information is contained in goods 1 and

10, while the information content decreases, with goods 5 and 6 being the least informative. Therefore, a polarization estimate strictly between 0.5 and 1 should be expected.

We then use the simulated data and the estimation algorithm to compute both mean and median polarization. The results are as follows:

| Simulation | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Mean Polarization | 0.5 | 0.5 | 1 | 0.704 |
| Median Polarization | 0.5 | 0.5 | 1 | 0.704 |

Table 2: Estimated polarization for the different simulations

First, we can see that the estimation is able to recover the theoretical polarization values for the first three simulations. In addition, we now have a reference for how to interpret the polarization estimates from the true data.

## *B. Robustness Polarization*

Here we present the plots for the polarization estimates that were not presented in the main part. These are the remaining estimates for the expenditure based specification as well as all the results for the specification based on household income. We also present all the results using quantities instead of expenditures to quantify consumption. Possible differences in polarization between the two grouping variables reveal additional information. If for instance the specification using expenditures instead of quantities shows a higher level of polarization for the same level of product aggregation, this suggests that more expensive products are more polarized than cheaper products.

### B.1. Households grouped by expenditures

For all three levels of product aggregation presented here, the estimated polarization is higher when we use expenditures to quantify purchases instead of the number of items bought. This suggests that more expensive items are more polarized than cheaper items. Similar to Bertrand and Kamenica (2023), we find that polarization is relatively stable over time and does not change much from 2004 to 2016.

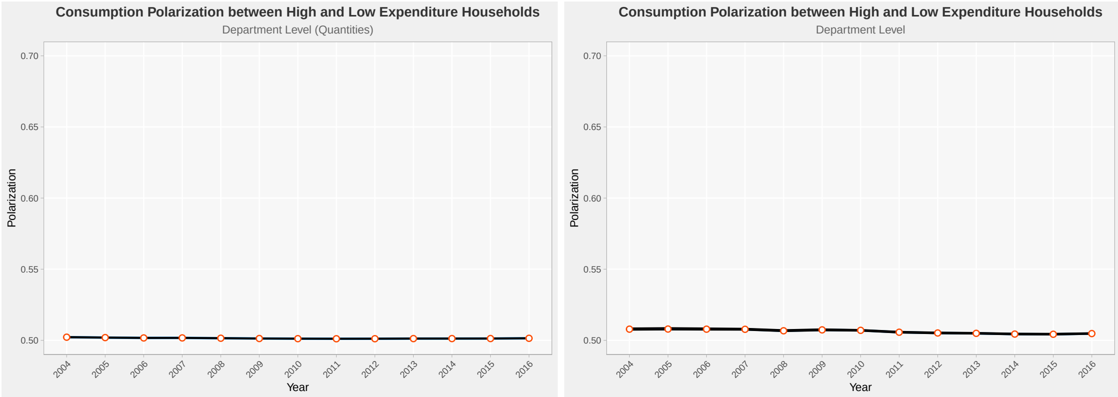



Figure 6: Both plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro, and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product department level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

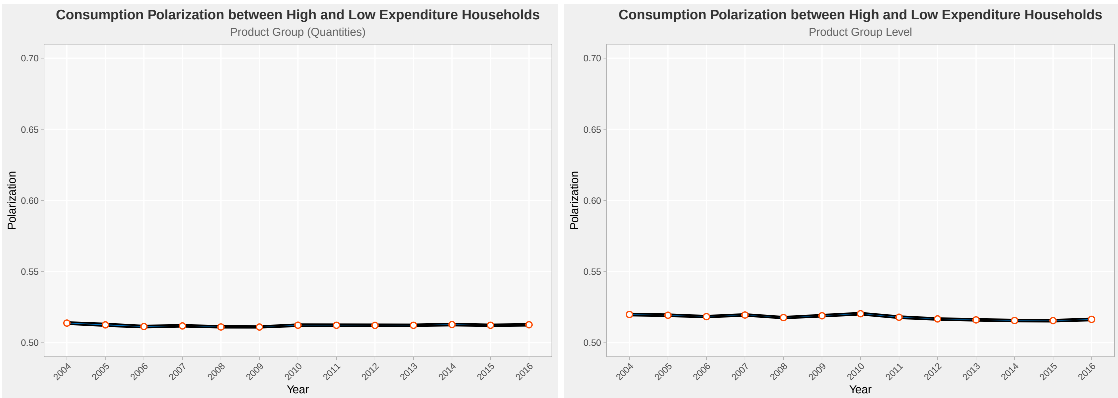



Figure 7: Both plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro, and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product group level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

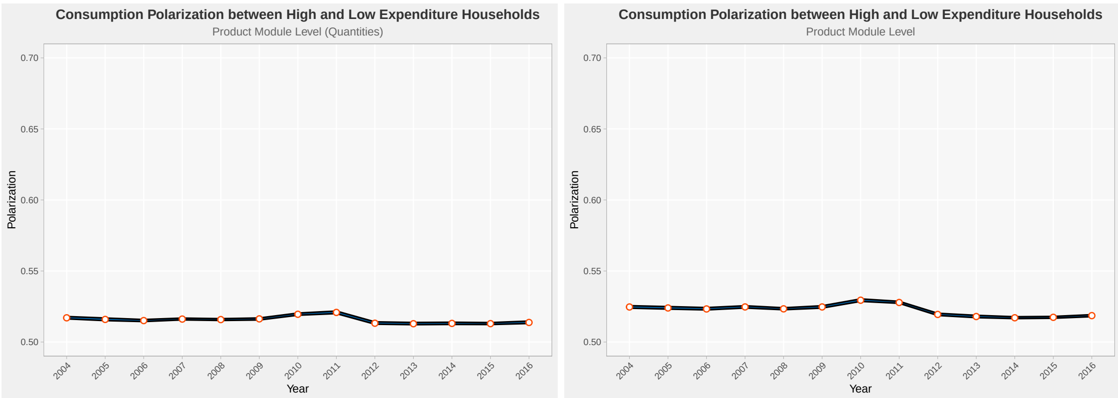



Figure 8: Both plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro, and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product module level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

### B.2. Households grouped by Income Brackets

Here we provide the results for the alternative household grouping. Instead of using the top and bottom 20% of the consumption expenditure distribution, we use the income variable available within the dataset. We define the high-income (low-income) group as the households belonging to the top (bottom) 20% of the income distribution. Due to the fact that households provide information about the income they received 2 years prior, the estimation is limited to the time period from 2004 to 2014 in this case. When comparing results to those obtained from the baseline specifications, the main conclusions do not change significantly. The estimated polarization levels are relatively similar, although overall slightly lower for the income-based grouping. All other differences are negligible and do not exhibit any conceivable pattern.

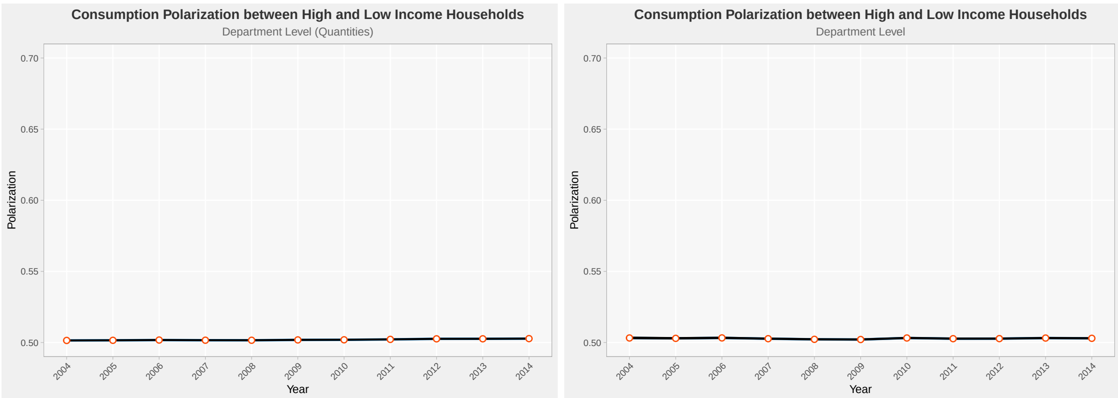



Figure 9: Both plots display polarization estimates between the top 20% and bottom 20% of the income distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro, and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product department level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.


Figure 10: Both plots display polarization estimates between the top 20% and bottom 20% of the income distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro, and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product group level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

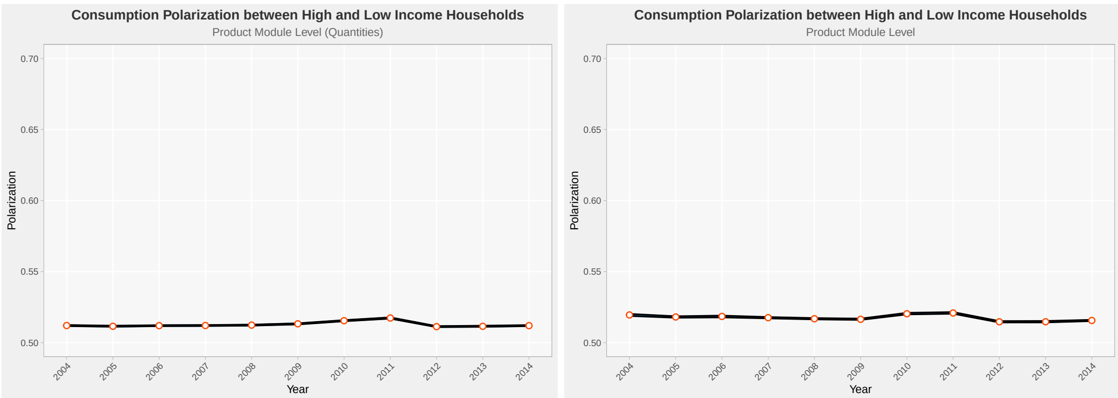



Figure 11: Both plots display polarization estimates between the top 20% and bottom 20% of the income distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro, and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product module level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

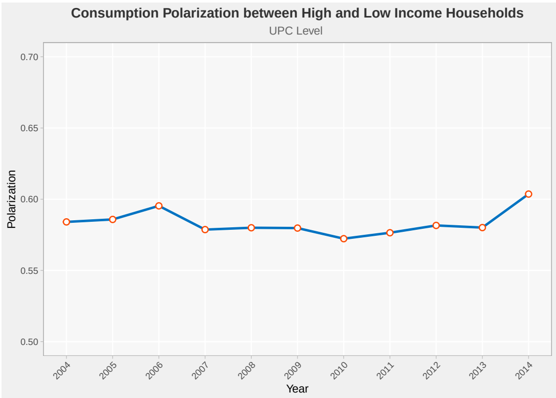



Figure 12: The plot displays polarization estimates between the top 20% and bottom 20% of the income distribution. The estimates are based on the within department estimates. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

*C. Aggregation of Polarization Measures*

Due to computational constraints, we are not able to estimate polarization at the UPC level directly for all products. Instead, we estimate polarization within each department separately. We then use these 10 polarization estimates to calculate aggregate polarization. The idea behind the aggregation is the following: Polarization, in our context, is defined as the probability of correctly guessing group membership from observing one random dollar of spending. Let $P(x)$ denote this probability. Additionally, let $P(x|y_i)$ denote the probability of guessing correctly, conditional on the purchase being made from department $y_i$, and let $P(y_i)$ denote the probability that a purchase is made from department $y_i$. Then, by the law of total probability, we have:

$$P(x) = \sum_{y_i} P(x|y_i)P(y_i)$$

where $P(x|y_i)$ is the within-department polarization, and $P(y_i)$ is the probability of purchasing a product from department $y_i$, which is estimated as a byproduct of estimating polarization at the department level.

*D. Polarization within Department*

The within-product department polarization estimates reveal a significant degree of heterogeneity between the different departments. While the observed level of polarization for most of the product departments is still below or around a value of 0.6 and, therefore, still not too far away from the results obtained for higher levels of product aggregation, we can see that 3 departments stand out as being significantly more polarized. These departments are Non-Food Grocery, Alcohol and General Merchandise. The departments with the lowest average level of polarization are Fresh Produce, Dry Grocery and Packaged Meat.

Almost none of the departments show signs of a time trend; only for General Merchandise there seems to be a trend toward higher levels of polarization. Additionally, we can see that for some of the departments, polarization is more volatile over time than at the aggregate level. The most volatile departments are Non-Food Grocery and Alcohol, while Health and Beauty Aids shows the lowest level of volatility. When we compare the results obtained from using income as the grouping variable, we can see that polarization levels are higher for the baseline grouping. While there are quantitative differences in the results, qualitatively there is no significant difference between the results obtained for the two different grouping variables.

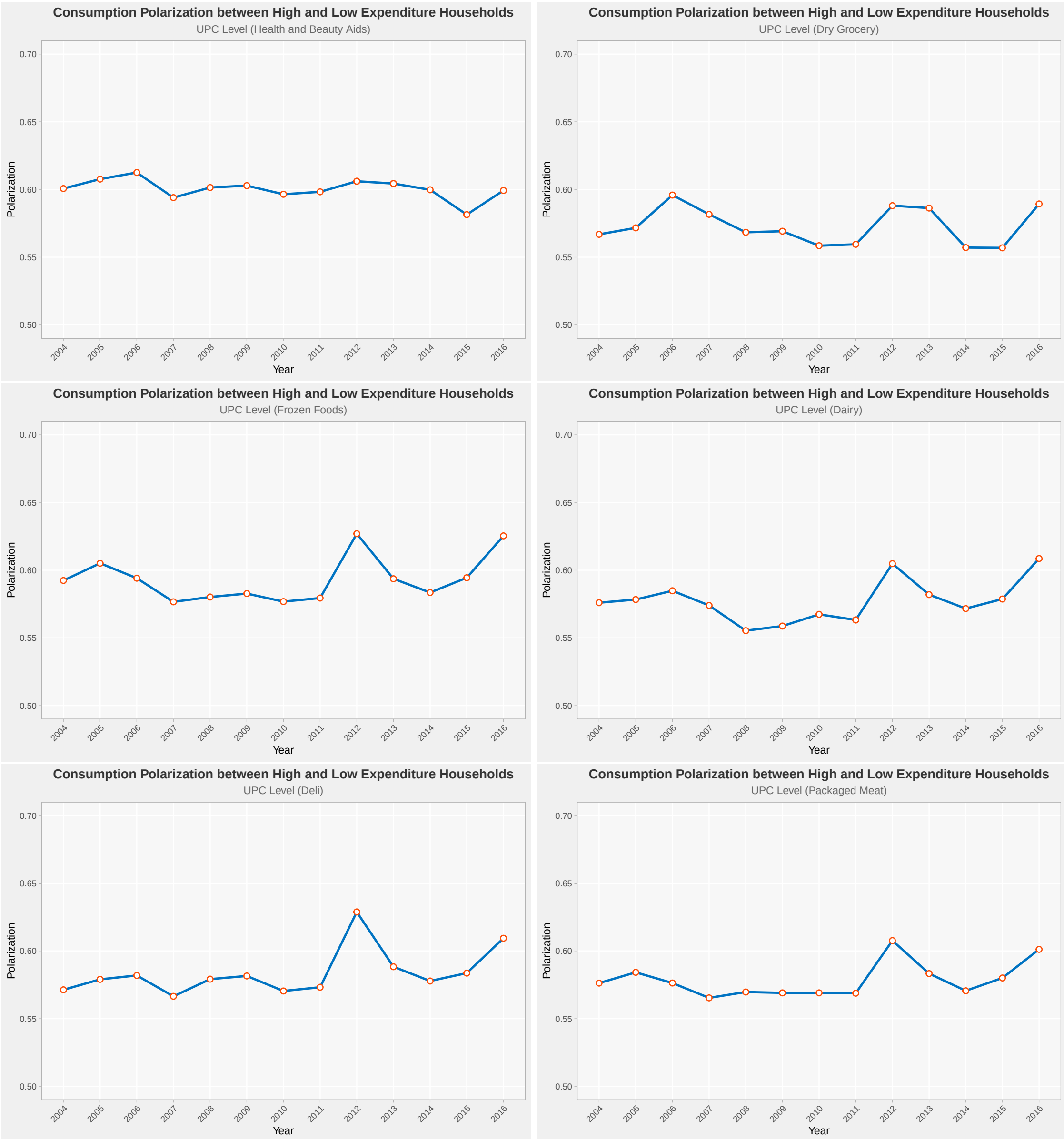



Figure 13: The plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

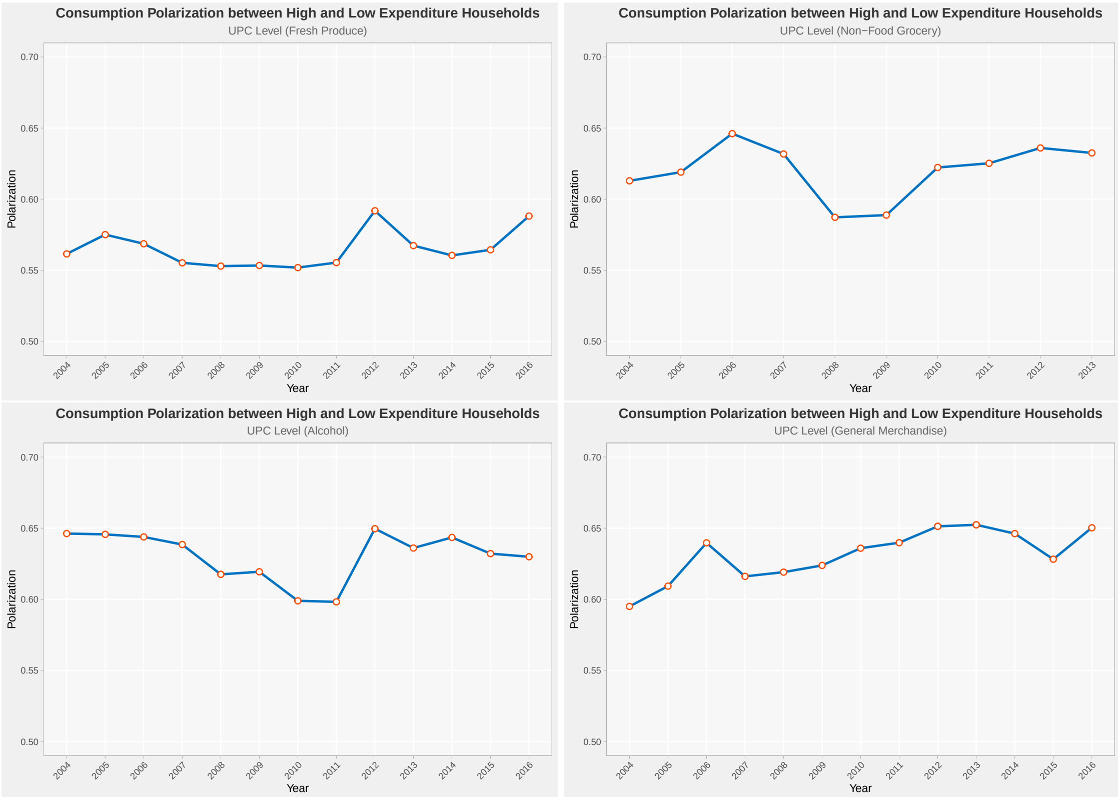



Figure 14: The plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

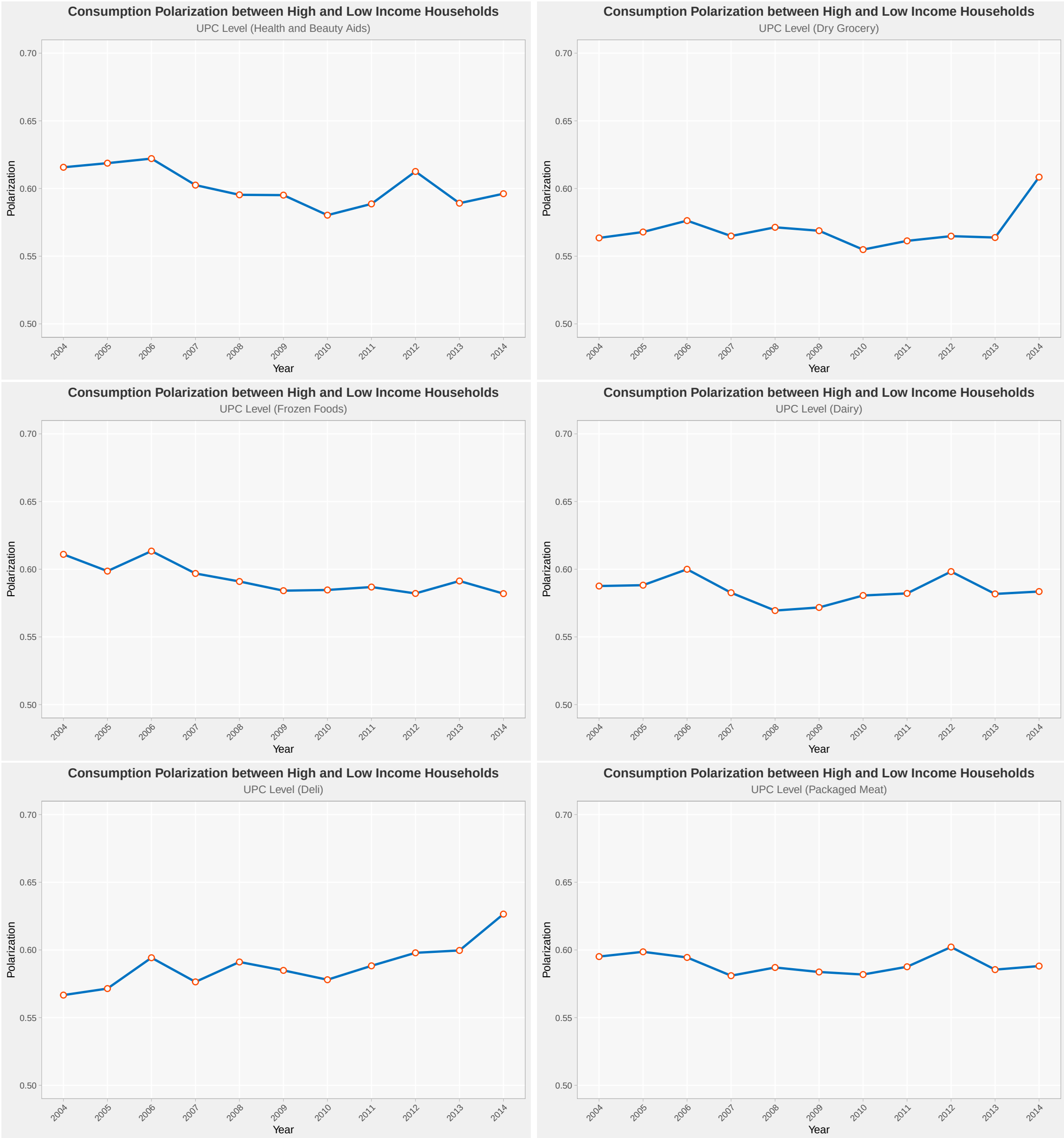



Figure 15: The plots display polarization estimates between the top 20% and bottom 20% of the income distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

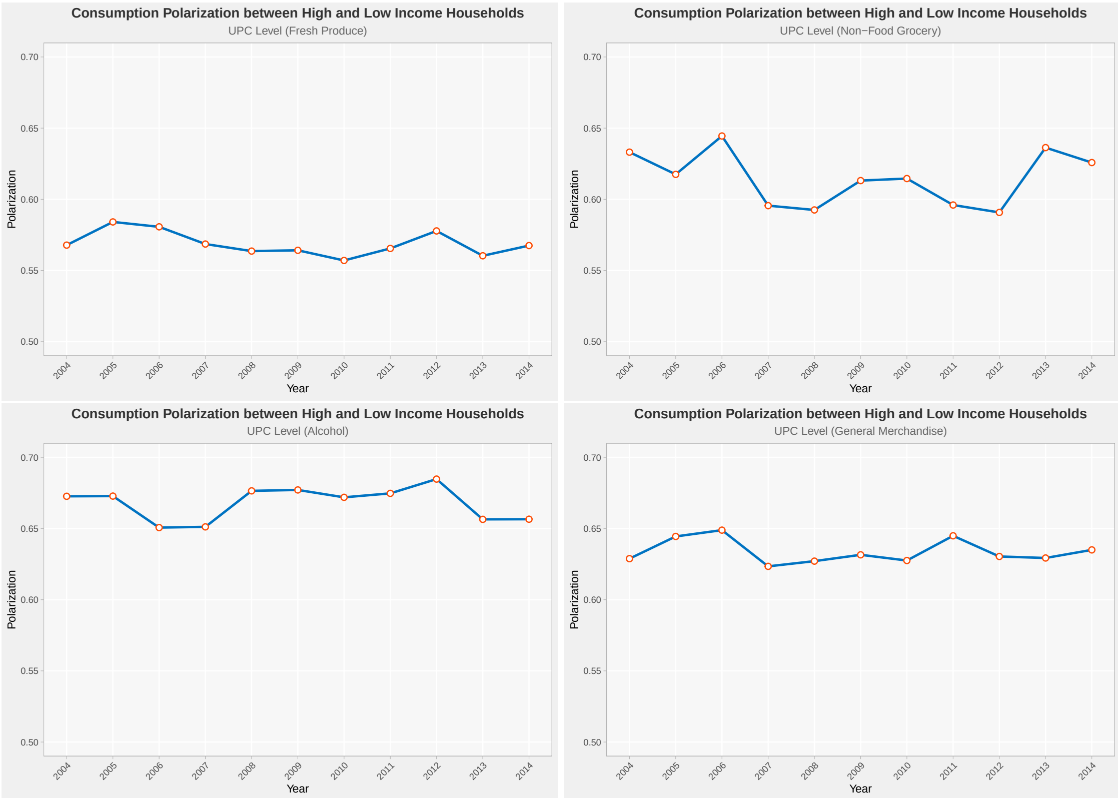



Figure 16: The plots display polarization estimates between the top 20% and bottom 20% of the income distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

*E. Robustness Persistence*

This section contains the plots from the persistence section that are not shown within the main body, as well as some additional robustness checks and persistence in terms of purchased UPCs. Before discussing the additional results, we define the measure for persistence of the consumed UPCs. Define $\mathcal{U}_{i,j}$ as the set of UPCs purchased by household $i$ in year $j$. Then the measure of overlap is given by:

$$O^{UPC}_{i,j+1} = \frac{|\mathcal{U}_{i,j} \cap \mathcal{U}_{i,j+1}|}{|\mathcal{U}_{i,j}|}$$

The left plot in Figure 17 shows the estimates of persistence of purchased UPCs. It is clearly evident that persistence is substantially lower in terms of UPCs than in terms of expenditures. This suggests that products with higher expenditure shares are more persistent than products with a relatively low expenditure share. The plot on the right shows estimates based on the expenditure measure. It presents both the persistence estimate using all available data as well as an additional estimate only including those products that are available within the US market in both years considered. Since there is no substantial visible difference between the plotted estimates, we can conclude that product exit does not contribute to low persistence in a meaningful way.

Figure 18 shows the baseline polarization estimates for two groups of households. One group consists of all households that experience a change in income bracket, and the other group consists of the ones that do not. Since there is no visible difference between the average polarization estimate in both groups, we can conclude that basket persistence is driven by factors other than income changes. Changing perspectives, this shows that even in the absence of income changes, consumption baskets change significantly from year to year.

The plot on the left in Figure 19 shows the histogram for the UPC-based measure, and the one on the right shows the histogram of household-level persistence within consumption expenditures, where, in addition to the projection factors, we use total household consumption expenditures as a weight. The first histogram

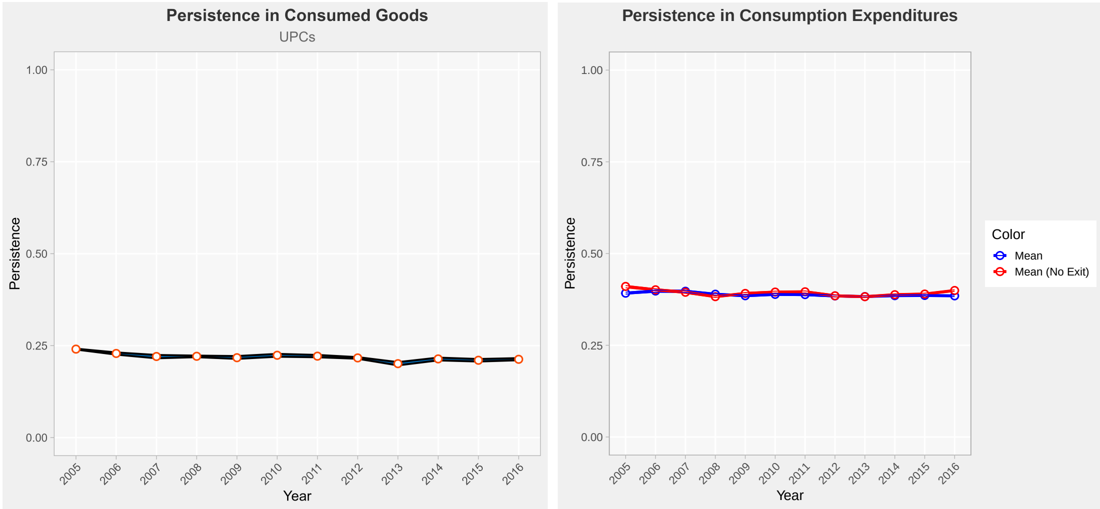



Figure 17: Persistence estimates. The figure displays the estimates for the expenditure based measure. No Exit refers to the estimates where all products are excluded that are not available to buy somewhere in the US in both years considered. All confidence bands are 95% and not visible because they are very narrow compared to the scale of the y-axis which is kept identical for all plots in this section.

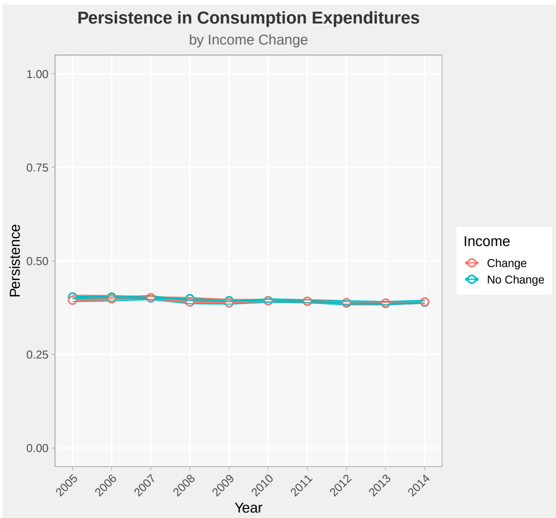



Figure 18: Persistence estimates for households that experience a change in income and those that do not. All confidence bands are 95% and not visable because they are very narrow compared to the scale of the y-axis which is kept identical for all plots in this section.

shows that, also within the UPC-based measure, there is substantial heterogeneity. It also clearly shows that there are almost no households with a persistence of more than 50% in terms of UPCs.

This histogram on the right is reweighted to give a higher weight to households that consume more to get a better sense how important low levels of persistence actually are in the overall economy. As we can see from the reweighted histogram, most of the mass is still below 0.5. This implies that substantial parts of overall consumption expenditures are made by households with low levels of persistence.

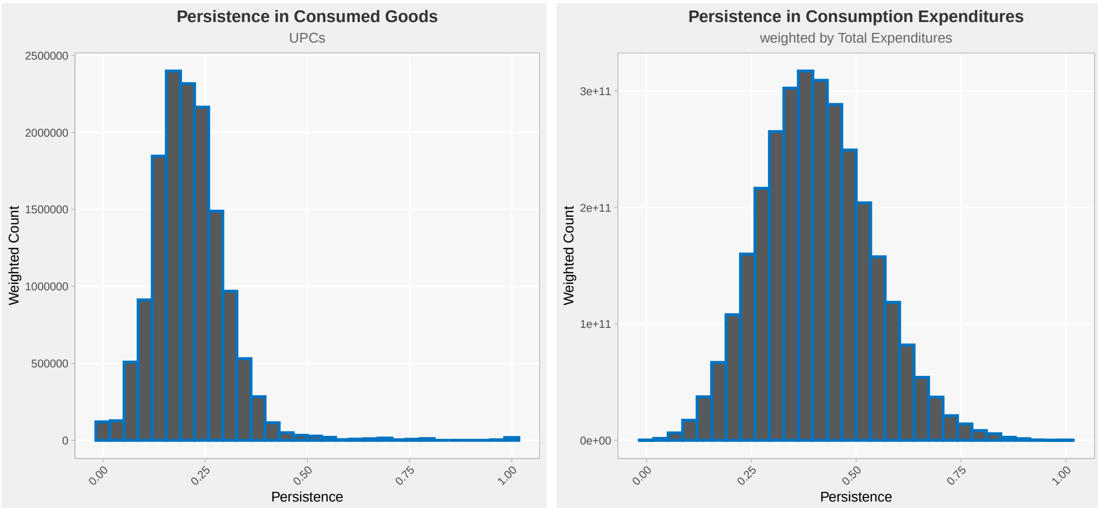



Figure 19: Histograms of persistence estimates. The figure shows a histogram of the persistence estimates for the expenditure based measure weighted by the projection factor as well as the households expenditures.

Finally, Figure 20 shows densities for the persistence measures split by the number of unique UPCs within the basket. One can see from the plot that the variance of persistence decreases as the number of consumed products increases. This suggests that persistence behaves as if it were to converge to its mean value as the number of consumed UPCs tends to infinity, or put another way, persistence becomes more stable as the number of products within the basket increases. One possible explanation for this kind of behavior would be that each product is roughly equally likely to be dropped from the basket. Then, as the number of consumed products increases, persistence would, by the law of large numbers, converge to the likelihood of a product being dropped from the basket.

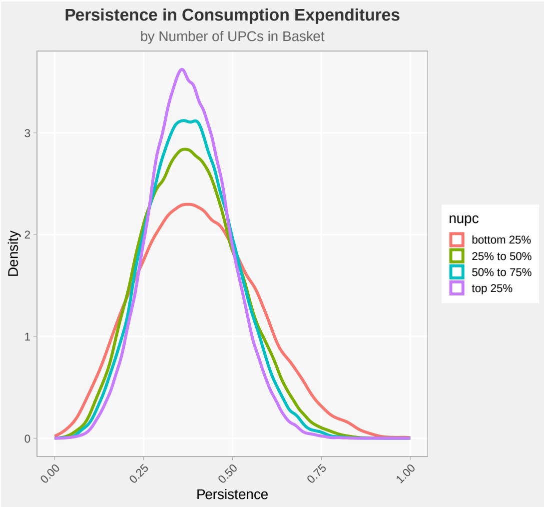
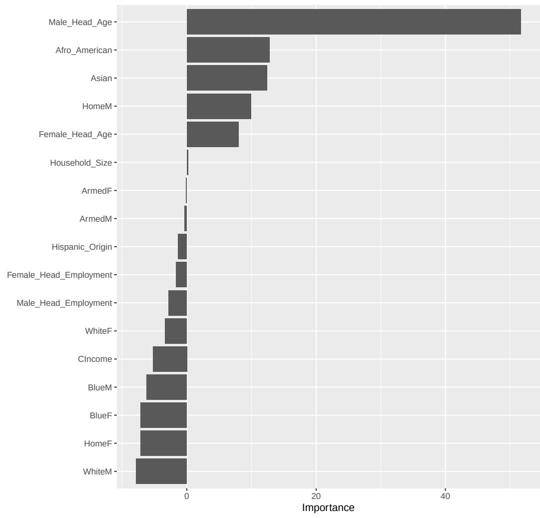


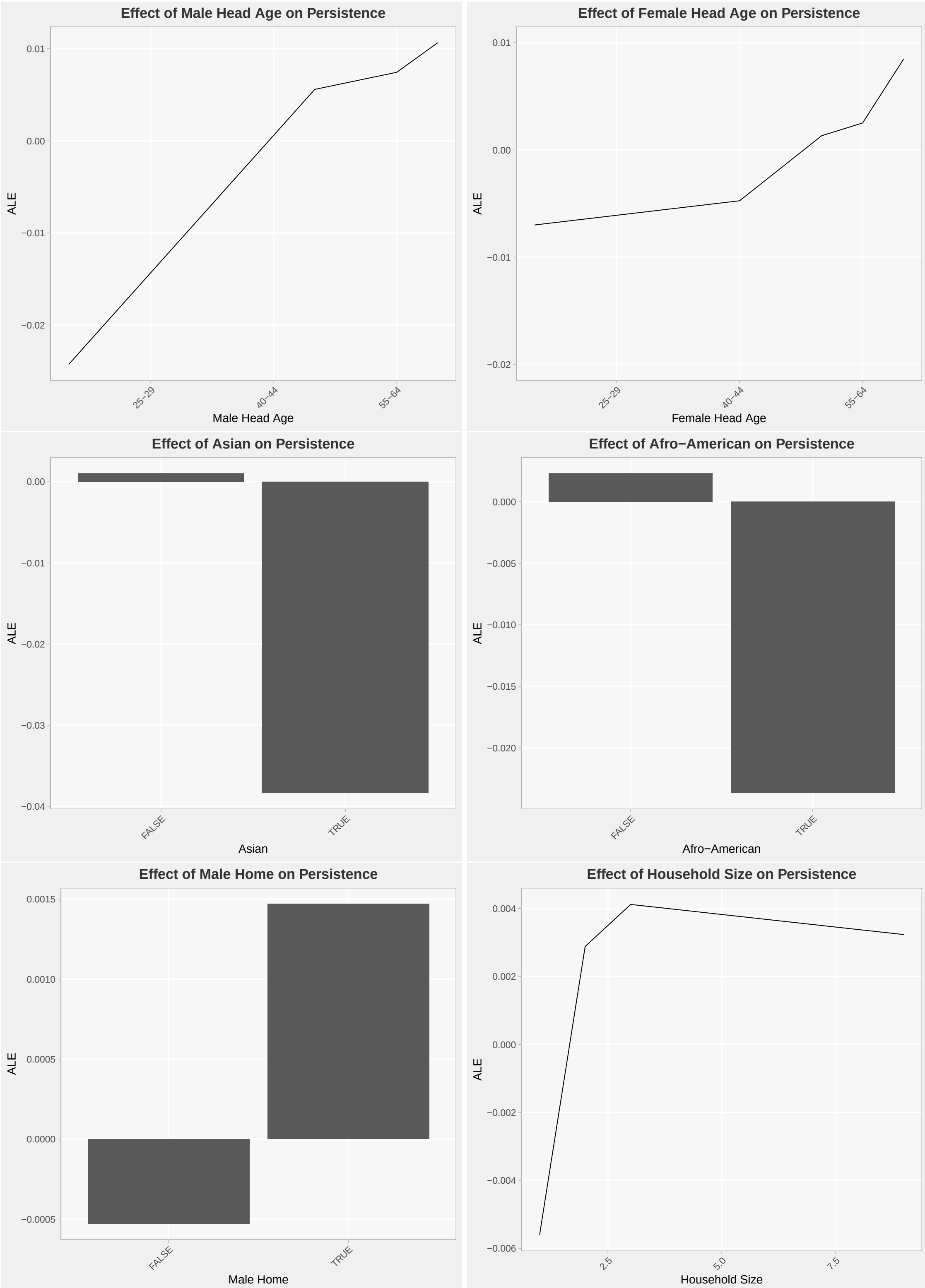

Figure 20: Densities of persistence estimates. The figure shows densities of the persistence estimates for the expenditure based measure where the households are split into 4 groups according to the number of UPCs in their consumption basket.

*F. Random Forest Analysis of Basket Heterogeneity*

Here we provide the results for the random forest analysis. The main idea behind a random forest is to identify the variable with the highest explanatory power. The results of the analysis can be visualized using a variable importance plot.




By far, the most important contribution comes from the age of the male household head. Additionally, the racial background, the age of the female head, as well as whether the male head is employed or non-employed, are identified as contributing to the heterogeneity within persistence. Variables like the size of the household, as well as income, offer little to no value in explaining persistence. Now that we have identified the variables that have the most explanatory power, the next step is to quantify the impact on persistence as well as the direction of the effect. To do so, we will look at the partial correlations between the explanatory variables and basket persistence. Since the assumption of zero correlation between the household characteristics is unlikely to be met, we use accumulated local effects (ALE) instead of partial dependence plots for the analysis.


Effect of Male Head Age on Persistence
0.01
0.00
−0.01
−0.02
ALE
25−29
40−44
55−64
Male Head Age
Effect of Female Head Age on Persistence
0.01
0.00
−0.01
−0.02
ALE
25−29
40−44
55−64
Female Head Age
Effect of Asian on Persistence
0.00
−0.01
−0.02
−0.03
−0.04
ALE
FALSE
TRUE
Asian
Effect of Afro−American on Persistence
0.000
−0.005
−0.010
−0.015
−0.020
ALE
FALSE
TRUE
Afro−American
Effect of Male Home on Persistence
0.0015
0.0010
0.0005
0.0000
−0.0005
ALE
FALSE
TRUE
Male Home
Effect of Household Size on Persistence
0.004
0.002
0.000
−0.002
−0.004
−0.006
ALE
2.5
5.0
7.5
Household Size

When examining the size of the ALEs and the very low $R^2$ value of approximately 0.06 for the random forest model, it becomes evident that the explanatory power of the considered variables is negligible. Hence, we must conclude that the observed heterogeneity in persistence remains latent. Regarding the effects, we observe that basket persistence increases with the age of both the male and the female head of the household. This may indicate that households, over their lifetime, become more stable in the kinds of products they prefer, possibly because they discover their own tastes over time or become more familiar with the products available in the market. Since we control for household size, we can be sure that this effect is not caused by changes in household size over time. Being of Asian or Afro-American descent has a negative effect on basket persistence. Employment of the male head is negatively correlated with persistence. For household size, persistence increases when moving from a single-person household to one with two members; beyond that, further increases in household size are associated with a gradual decrease in persistence.