Three Essays in Macroeconomics

Daniel Runge

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

University of Mannheim
May, 2025

Advisor: Krzysztof Pytka

Dekan:	Prof. Dr. Thomas Tröger
Referent:	Prof. Krzysztof Pytka, Ph.D.
Koreferent:	Prof. Dr. Moritz Kuhn
Vorsitzender der Disputation:	Prof. Dr. Antonio Ciccone
Tag der Disputation:	17.09.2025

Contents

I. Acknowledgments

I am very grateful for the advice, guidance and encouragement of my supervisor and coauthor Krzysztof Pytka. Without his support it would not have been possible to complete this work. I thank Antonio Ciccone and Moritz Kuhn for becoming part of my dissertation comittee.

I am grateful for the opportunity to study at the the Graduate School of Economics and Social Sciences at the University of Mannheim and for the encouraging and positive environment provided by the University. My research has benefited a lot from the fruitful and critical discussions within the CDSE Seminar as well as the Mannheim Macro Seminar. I especially want to thank David Koll, Antonio Ciccone and Konrad Stahl for their valuable feedback on my research.

I thank the GESS for the support during years of studies. I am also very thankful that Krzysztof Pytka provided me with the opportunity to work with the Kilts-NielsenIQ Consumer Panel for this dissertation and that NielsenIQ approved the project.

Finally, I am also very thankful for the support and encouragement of my family and friend, especially Raphaela and Lara.

II. Disclaimer

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.

III. Introduction

This work contains three essays on Macroeconomics and the implications of dispersion in paid prices as well as the consumption choices of households. All three make use of the Kilts-NielsenIQ Consumer Panel, which documents households' consumption choices and prices paid at the individual product level. This allows for an accurate measurement of, for example, paid prices, but also inflation rates directly at the household level. Additionally, it opens the possibility of comparing consumption baskets of different households or the same households over time at a level of granularity so far absent in the literature. Measuring directly at the household level is important to draw valid conclusions about either inequality or welfare.

I want to stress that all three essays are not only connected by making use of the same data source, but are also inherently linked by their content. The first and third essays are the result of unanswered questions that arose during the work on the second one. I want to use this introduction to give the reader a sense of how exactly the three are connected. In order to do so, I will briefly highlight the main content of each essay, before highlighting the linkages and findings in more detail.

The first of the essays deals, with the question of how similar the consumption choices of households are along the income distribution. In addition, it examines how much the composition of the consumption basket of given households varies over time. The second essay then examines the effect of monetary policy shocks on households' choices. In particular, it focuses on the effect on paid prices, relative prices, and the quality of households' consumption baskets. The third essay examines the validity of the idea of proxying product quality using goods prices from a dataset that features a high degree of sparsity in the presence of price search frictions.

The first of the essays lays the groundwork. It shows to what degree high- and low-income households consume the same goods and how stable consumption baskets are over time. Both of these findings inform the analysis conducted in the second essay. For instance, it is crucial to first know how much overlap exists in the consumption choices between different groups of households to judge the

relevance of results on changes in relative prices, i.e. the percentage difference in prices for identical goods. In addition, when constructing inflation indices directly at the household level, it is crucial to know how high the turnover is in households' consumption baskets, as there is no paid price observed for products that are no longer purchased by a given household. Finally, the last essay asks whether observed differences in product quality, measured using paid prices, could result artificially from the interaction of price search and sparsity. The issue of product quality is important in this context as the first essay presents some evidence indicating that there is at most a weak product quality ladder and within the second essay product quality is used to understand if and how household change to composition of their consumption basket in response to a monetary policy shock.

Chapter 1 written jointly with Krzysztof Pytka 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.

Chapter 2 reexamines the impact of monetary policy shocks on prices, acknowledging that paid and posted prices are not the same. I use consumer microdata from the Kilts-NielsenIQ Consumer Panel jointly with identified monetary policy shocks to analyze the effect of monetary policy. I find that an undecomposed contractionary shock lowers paid prices, but a pure monetary policy shock unexpectedly increases them. Households adjust their price search behavior differently depending on their income level, with low-income households increasing their search effort relative to the high-income group. While there is no evidence of the overall product quality of households' consumption baskets changing in response to a shock, within-department analysis reveals that quality adjustments take place. The study highlights the importance of considering paid prices and household heterogeneity in monetary policy analysis.

Chapter 3 explores how, when quality is measured using a product's price,

data sparsity and search frictions can lead to artificial quality differences between products. Using a search-theoretical model, I demonstrate that even when products are identical, sparsity combined with price search frictions can create spurious quality ladders. Furthermore, it can lead to falsely concluding that higher-income households consume higher-quality goods. In addition, using data from the Kilts-NielsenIQ Consumer Panel (KNCP), I find that quality differences appear smaller for products with more purchases, suggesting that indeed small sample bias plays a role in the quality estimation.

Chapter 1

Looking into the Consumption Black Box: Evidence from Scanner Data

Joint work with Krzysztof Pytka[1]

[1]We are thankful for comments made by Fernando Alvarez, Łukasz Drozd, Miklós Koren, Dirk Krueger, Luigi Paciello, Jesse Shapiro, Michelle Sovinsky, and Philipp Wangner. Lion Szlagowski's stellar research assistance was invaluable to this work. All errors are ours.

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.

I. Introduction

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 multi-

nomial 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 influ-

enced 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. 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, 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).[2] 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.

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

[2]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.

stance on consumption sorting across the income distribution, as seen in Nord (2023), Becker (2024), Sangani (2024), 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.

II. 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.

To construct our sample, we exclude all items without a barcode (labeled as "magnet" within the dataset). The reason for this exclusion is that these products are not reported by all households, and including those products would require restricting the sample to only those households reporting these products, thereby severely reducing the overall sample size.

III. 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

This section will compare the consumption choices of high- and low-income households, as previous research has identified large differences in shopping behavior along this dimension. It is important to know the shared amount of consumption between these two groups, for instance, when considering the search intensities a retailer should expect for a given product. These search intensities are one key determinant of a retailer's market power.

An algorithm suitable for this kind of comparison must satisfy two criteria: First of all, it needs to be able to compare groups in a high-dimensional space while still being computationally feasible, as well as provide clearly interpretable results. Second, we need to take some sparsity within the dataset itself into account, therefore the algorithm needs to be able to perform the estimation without producing artificial results. Gentzkow, Shapiro and Taddy (2019b) have introduced a methodology to estimate polarization that satisfies these criteria. Overall, our analysis fits into the spirit of the generative model approach outlined in Gentzkow et al. (2019b): The main idea is to infer properties of the data-generating process from the data we observe. In the given case, this implies understanding the households themselves as consumption-generating processes and inferring from the observed choices the properties of these consumption-generating processes. The group comparison is then performed by comparing the two estimated processes.

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:

$$\mathbf{c}_{i,t} \sim \text{Multinomial}(m_{i,t}, \mathbf{q}_t^{P(i)}(\mathbf{x}_{i,t})) \tag{1.1}$$

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.[3] 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:

$$\begin{aligned} q_{jt}^{P(i)}(\mathbf{x}_{i,t}) &= \frac{\mathbf{e}^{\mathbf{u}_{i,j,t}}}{\sum_l \mathbf{e}^{\mathbf{u}_{i,l,t}}} \\ u_{i,j,t} &= \alpha_{j,t} + \mathbf{x}'_{i,t}\boldsymbol{\gamma_{j,t}} + \varphi_{j,t}\mathbf{1}_{i\in H_t} \end{aligned} \tag{1.2}$$

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

[3] 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.

household characteristics influencing choice probabilities are those included in $\mathbf{x}_{i,t}$, besides 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 (1.2).[4]

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 polarization measure.[5] 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:

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

Let, H_t be the set of high-income households and L_t the set of low-income households at time t. Then, *average polarization* is given by:

$$\bar{\pi}_t = \frac{1}{|H_t \cup L_t|} \sum_{i \in |H_t \cup L_t|} \pi_t(x_{it}). \tag{1.3}$$

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

[4] 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.

[5] 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.

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. 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 (1.2) 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 documented by us in 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.[6] To mitigate this problem, we regularize $\varphi_{j,t}$ in Equation (1.2), 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 mostly 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.[7]

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. To compute standard errors we adhere to the process detailed in Gentzkow, Shapiro and Taddy (2019b). Due to computational constraints, we do not report confidence bands for the polarization measure at the barcode level.[8] For the same reason, we conducted the polarization analysis at the most granular level—the barcode—by

[6] We also examine transaction frequency on an annual basis (Runge, 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.

[7] 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.

[8] 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.

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.

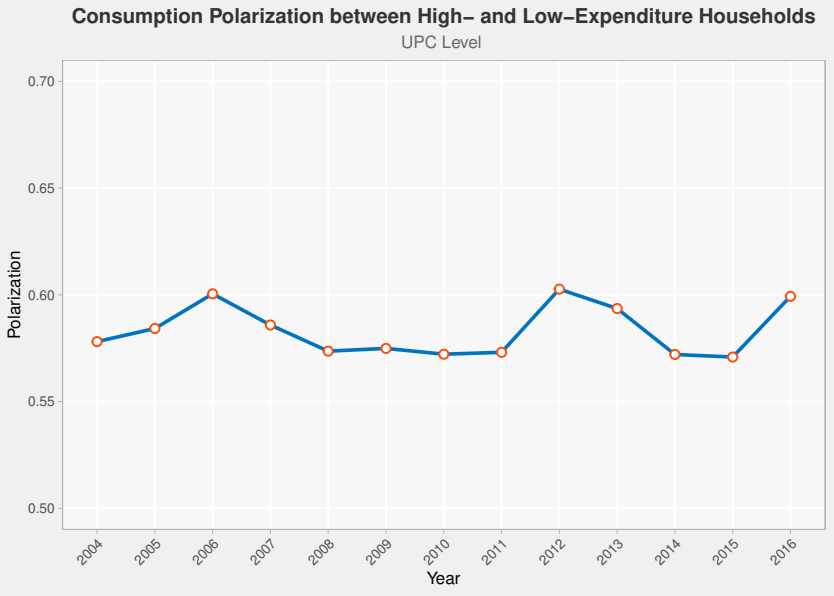


Figure 1.1: The plot displays polarization estimates between the top 20% and bottom 20% of the consumption expenditures distribution. The estimates are based on the within-department estimates. No confidence intervals are provided. All products are defined at UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

Figure 1.1 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 how it was allocated, the probability of correctly classifying the spender is:*

1. *58.3% when the* ***barcode*** *of the purchased product is observed,*
2. *52.2% when the* ***product module*** *is observed,*

3. *51.7% when the* ***product group*** *is observed,*

4. *50.6% when the* ***product department*** *is observed.*

These results indicate that income-based consumption sorting is weak, as product choices across income groups exhibit substantial overlap even at detailed classification levels.

We can observe some features that are common to the results for all of the estimated specifications. First of all, polarization is relatively stable and shows no trend. By construction, polarization decreases as the level of aggregation increases. At product department level, we can clearly see that polarization is very low, with a maximum value of about 0.51. Given that there are only 10 product departments in the data used for the analysis and households are expected to consume products from all of those departments, this result aligns with our expectations and can be seen as a sanity check for the procedure we used. Even in the analysis at product group or module level, polarization is still quite low and stays below 0.52 and 0.53, respectively. This suggests that at these levels, it is nearly impossible to differentiate between the two income groups based on just their consumption choices. When comparing the specifications using quantities to the ones that use expenditures to measure consumption, it becomes clear that polarization is always slightly higher in the expenditure-based specifications. This indicates that products with a higher price are more polarized than products with a lower price.

At the UPC level, polarization stays below a value of 0.6 and there is no visible trend within the estimates, showing that consumption polarization has remained almost unchanged from 2004 to 2016.

The fact that we observe relatively low levels of polarization even at this granular level might appear to stand in stark contrast to Bertrand and Kamenica (2023), who find a probability of about 90% to infer the correct income group of a household based solely on consumption choices. Overall we view our results as complementary rather then conflicting. The difference in results can be traced mainly to differences in methodology. Besides different data sources, the measures to estimate polarisation work differently. Even though both show the probability of correctly guessing group membership, they differ significantly in the way this probability is computed and the information included. 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. The measure Bertrand and Kamenica (2023) use computes instead the probability based on the product with the highest predictive power. This implies that as soon as there is one product that shows a very high level of polarization, overall polarization is estimated to be very high, even though this product might only account for a negligible amount of overall household spending.In our approach the influence of such a product would average out instead.

All in all, we show that polarization within consumption is much lower than often suspected. At both the department and product module levels, it is nearly impossible to outguess a simple coin flip when it come to guessing whether a household belongs to the top or bottom 20% of the consumption expenditure distribution from observing just a random dollar of spending. Even at the UPC level, it remains very unlikely.

B. Individual Consumption Baskets: Constantly in Flux

After comparing consumption baskets between high- and low-income households, we will now turn to comparing the consumption choices of individual households over time. Roughly speaking, we want to answer the following question: given that we know a household is consuming a good in a given year, how likely is it that the same good will also be consumed in the next year? We will call the degree to which a household's consumption choices are stable over time "persistence". A sound empirical understanding of persistence is important in many different economic contexts. For example, it indicates the validity of concepts like product or brand loyalty. More importantly, persistence has direct implications for situations in which a firm or retailer wants to enter a preexisting market with a new product. If households are highly persistent in their choices, then incumbents have a big advantage and it will be very hard for the firm to attract customers for its new product. If instead, households are very impersistent, entering the market will be much easier and the advantage of the incumbents smaller since households switch purchased products regularly. In this case, the level of persistence will directly affect the severity of the market entry barriers faced by newly entering

firms or products. We utilize two distinct measures to assess persistence within the consumption basket, each with a slight variation in focus. The first measure quantifies the overlap in consumed UPCs between two successive years, the second measures the fraction of expenditures in the second year spent on products already purchased in the first year. While the first one shows how stable the collection of products consumed by an individual household is, the second measure reveals how relevant the goods still consumed are in terms of consumption expenditures. We will start by formally defining the UPC-based measure. Define $\mathcal{U}_{i,j}$ as the set of UPCs purchased by household i in year t. Then the first measure of overlap is given by:

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

With this measure, each product is assigned equal weight regardless of the quantity consumed and the price paid for the product. It therefore provides a first insight into the stability of the baskets and the extent to which a consistent array of products is consumed over an extended time frame, disregarding their relative importance.

In contrast, the second measure is based on the expenditures allocated to each of the products. Therefore, the relative size in terms of expenditures in the basket of the households is taken into account. To define the measure, denote by $e_{i,t}(j)$ the expenditures on the good with UPC j by household i in period t. Then the overlap measure is given by:

$$O_{i,t+1}^{E} = \frac{\sum_{j \in (\mathcal{U}_{i,t} \cap \mathcal{U}_{i,t+1})} e_{i,t+1}(j)}{\sum_{j \in \mathcal{U}_{i,t+1}} e_{i,t+1}(j)}$$

We will compute both measures at household level, standard errors as well as confidence bands will be computed using bootstrapping. The results for the expenditure-based measure 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:*

1. *38.8% for products defined at the* ***barcode*** *level,*

2. *59.5% for products defined at the* ***brand*** $\times$ ***module*** *level,*

3. *83.5% for products defined at the* ***module*** *level.*

Overall, the analysis presented here reveals a surprisingly low level of persistence. To analyze persistence over time, we compute yearly averages of the household-level persistence measures, using the projection factors provided in the dataset as weights. Persistence in terms of purchased UPCs is extremely low, with a high point of approximately 25%. In terms of expenditures, persistence is significantly higher, with around 40%. This suggests that the consumption of high-expenditure items is more stable than consumption of items with comparatively lower expenditure shares. While the UPC-based measure could notably be impacted by households experimenting with new products or making other one-time purchases, this influence should be much less pronounced in the expenditure-based measure. Overall, these findings indicate that consumption is highly irregular, with on average only 40% of consumption expenditures being spent on recurring products.

To ensure that this effect is not driven by products entering and exiting the market, we recalculate the expenditure-based measure using only those UPCs that are available in both periods somewhere within the US. The difference between the two measures is economically insignificant. Therefore, the relatively low persistence of consumption baskets must be driven by other factors.

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.

We also investigate whether persistence differs for households that experience a change in income compared to those that do not. To do this, we recompute

the mean persistence and split the households into two groups: one for households that change income brackets and one for those that do not. We can show that there is no economically significant difference between these two groups. Therefore, income changes are not a primary cause of the instability. Plots of the results for the specifications referred to above can be found in Appendix E.

So far, we have only considered the overlap in expenditures between two consecutive years. As next step, we check how persistence changes when we consider longer time horizons. To do this, we compute persistence in consumption expenditures over up to 4 years, meaning the fraction of expenditures in a given year spent on products already bought 4 years earlier. As can be seen from the left plot in Figure 1.2, as the time horizon increases, persistence drops. To better understand how to interpret the observed drop in persistence imagine two extreme scenarios. In the first one all products are equally likely to be dropped from the basket. Then in a given year 60% of expenditures are dropped and 40% remain in the basket in the next year. If we consider one year further then of those remaining 40% again 60% would be dropped and the observed level of persistence would be 0.16%. In the second scenario, the basket consists of a fixed spending of 40% on products that are never dropped and 60% on products the household experiments with and then drops. In this scenario persistence at the two year mark would again be 40%. The level of persistence we observe in the data for products bought 4 years ago is about half of the fraction spent on products already bought in the previous year. Notably, persistence at the 2-year horizon is higher than 40% of persistence at the 1-year horizon. Taken together, this suggests that some products are less likely to be dropped from the basket than others.

After showing that average persistence at the UPC level is very low in terms of purchased products as well as in terms of expenditures, we will now try to better understand which kinds of products replace the ones that are dropped. Therefore, we recompute the persistence measure at higher levels of product aggregation. We will consider persistence in expenditures at brand, product module and brand $\times$ module levels. As we can see in the right plot in Figure 1.2, persistence at the module level is with 80% much higher than at the UPC level (40%). The results for the brand $\times$ module level (about 58%) show that roughly half of the difference

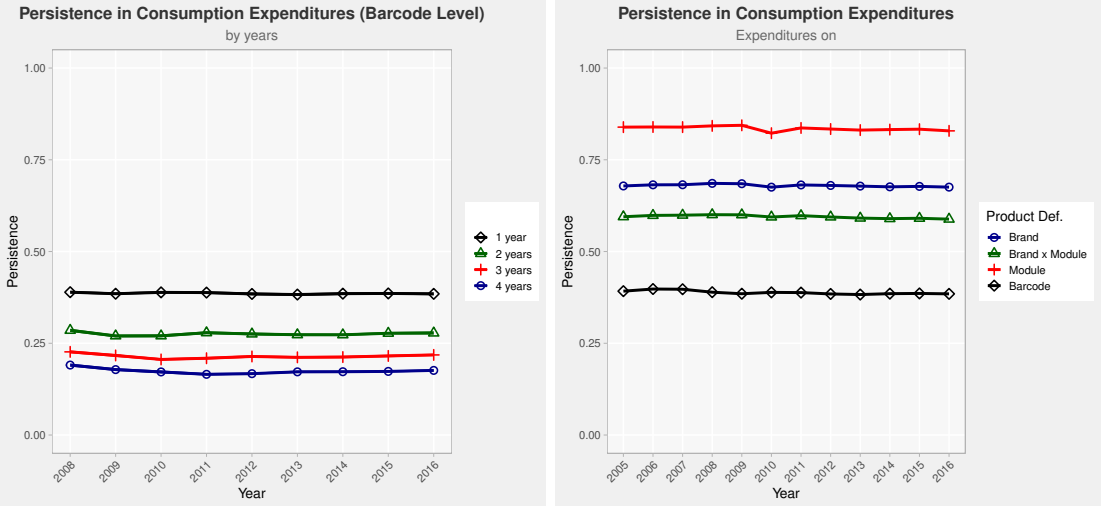


Figure 1.2: The plot on the left side shows persistence estimates at the barcode level for different time horizons and the plot on the right persistence estimates for different product definitions. All confidence bands are 95%. They are not visible in the plots since they are very narrow.

between module and UPC level is caused by product changes within the same brand, while the other half is caused by switching to products from a different brand within the same module. At brand level itself, persistence is with about 62% slightly higher than for the brand $\times$ module level, which captures the fact that a product might be replaced by one from the same brand but belonging to a different product module. Taken together, this can be seen as evidence that household preferences for types of products are quite stable, while the individual products that households consume change substantially.

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.

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.

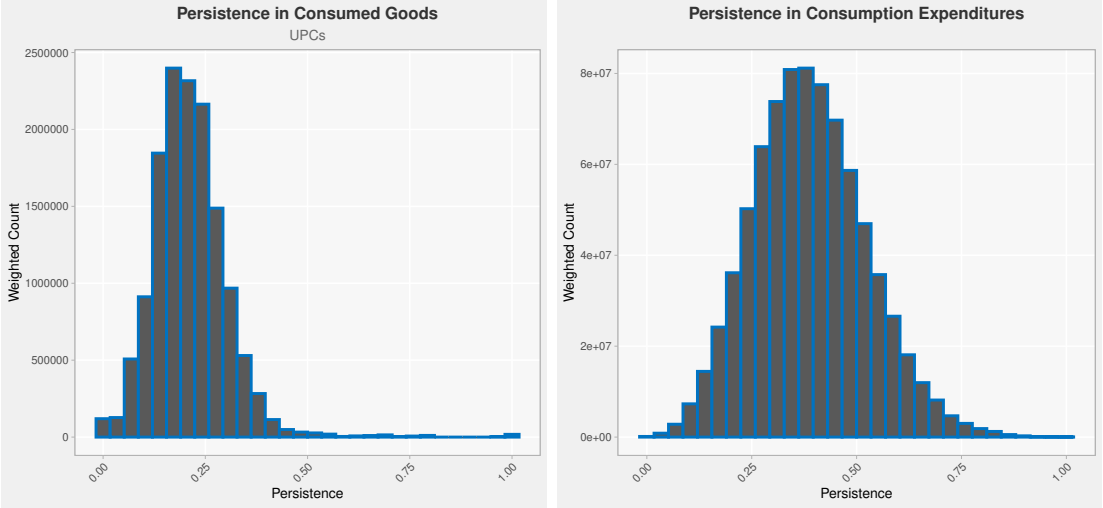


Figure 1.3: Histograms of persistence estimates. The figure on the left shows the histogram of persistence estimates using the UPC-based measure and the one on the right the expenditure-based one. In both figures the estimates are weighted using the projection factors.

Next, in our analysis we examine heterogeneity in individual (barcode level) persistence across households. To do so, we look at the distribution of the persistence measures using histograms. Figure 1.3 shows the histograms for the UPC and expenditure-based measures. One can observe that both measures of persistence are roughly symmetric around the mean, with relatively strong heterogeneity, which is more pronounced for the expenditure-based measure. The most striking feature of the distributions is that, for the UPC-count-based measure, almost all of the mass is below 50%, and for the expenditure-based measure, it is below 75%. This implies that there are almost no households for which, in terms of UPCs, the consumption basket changes by less than 50% from one year to the next. Even for the expenditure-based measure, we can see that more than 50% of the mass is below a level of persistence of 0.5, which implies that more than 50% of the households spent 50% or more of their expenditures in a given year on products

they newly added to their consumption basket. At the same time, the observed level of heterogeneity is striking.

To get a sense of how these findings transfer into aggregate consumption, we look again at the distribution of persistence measures, weighted by each household's total consumption expenditures as well as the projection factors. The results are vastly similar, and over 50% of aggregate consumption expenditures are accounted for by households that spent less than 50% of their expenditures on products already bought in the previous year. The histogram can be found in Appendix E. This clearly shows that most of aggregate non-durable consumption is accounted for by households with highly unstable consumption baskets.

We look at the distribution of the persistence measures, split by the number of unique products consumed by the household. The corresponding plot is shown in Appendix E. To ensure comparability, we use densities instead of histograms. The densities clearly show that the variance of the persistence measures decreases as the number of consumed products increases. This finding hints at a link between the number of consumed products and the stability of the basket, which is potentially caused by information frictions.

Finally, we trained a random forest to predict persistence using a set of sociodemographic 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.

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 im-

portant 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 for example 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).

IV. 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.[9]

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 prefer, we assume that all products are perfect substitutes (adjusted for prices).[10] 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.[11] This means that random 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

[9] 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.

[10] 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.

[11] 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.

of the shopping process for household i is summarized in Algorithm 1.[12]

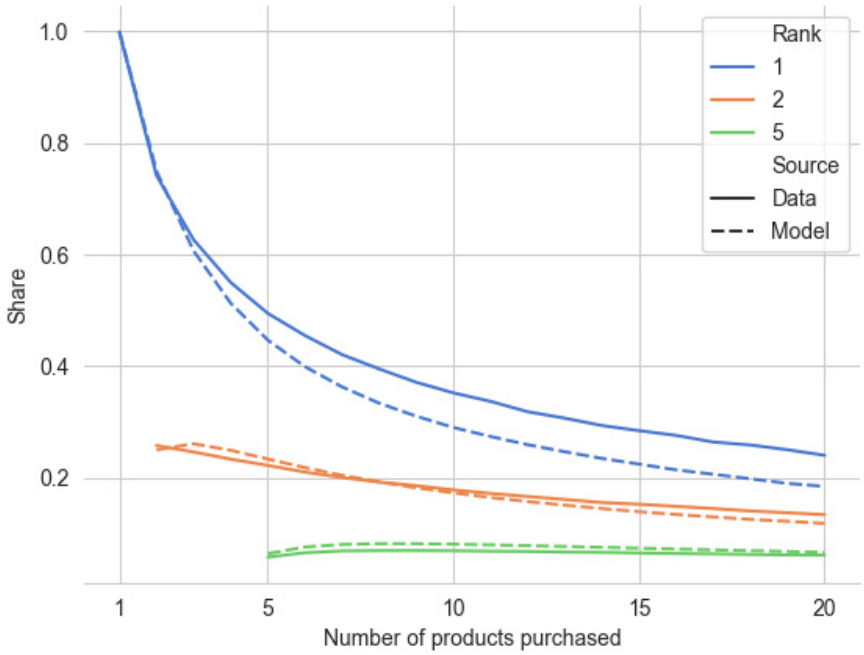
Algorithm 1 Shopping Spree Model

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.**

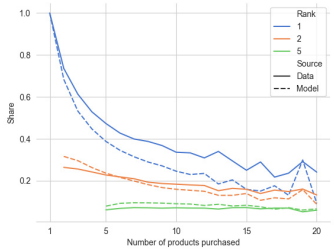
The model outlined here can be seen as a data-driven, high-dimensional extension of the classical model of impulsive customers proposed by Becker (1962). Compared to that model, we expand the consideration set to include all barcode-level products in the NielsenIQ universe, rather than just two as in 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.

Figure 1.4a illustrates the relationship between household spending shares on their first-, 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

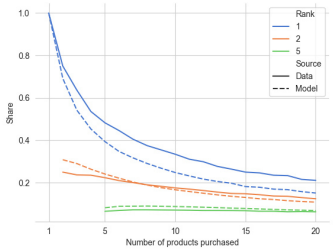
[12] Admittedly, it is possible for households to slightly violate their budget constraint, m_{it}. However, given the scale of annual expenditures and the spending on individual products, the magnitude of these violations is negligible.



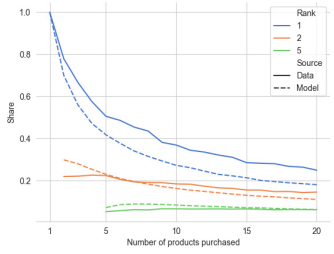
(a) Average Spendings on Different Ranked Items



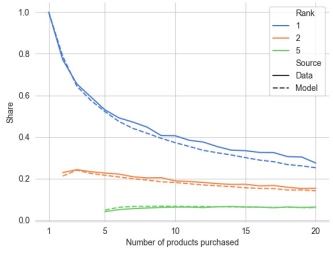
(b) Seafood Canned



(c) Cereal



(d) Yogurt



(e) Pet Food

Figure 1.4: Comparison of Average Spendings and Category Spendings on Different Ranked Items

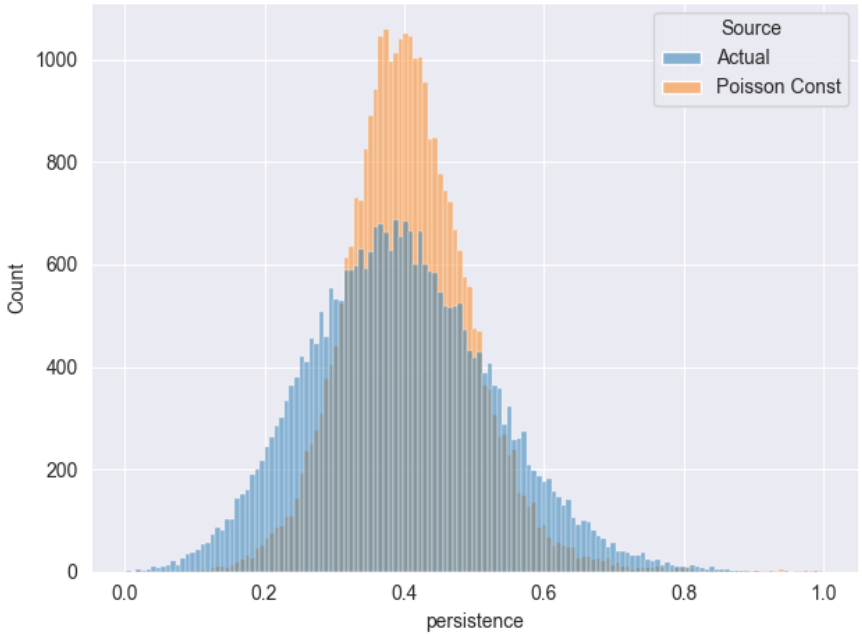


Figure 1.5: Household Consumption Persistence

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 1.4b to 1.4e 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 1.5. 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 1.20 confirms this, showing that households with fewer transactions exhibit greater dispersion in persistence.

V. 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, 2025) proposes a simple yet powerful model experiment in which search frictions, combined with data sparsity, can generate a spurious quality ladder.

Bibliography

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

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

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.

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.

Bornstein, Gideon (2021) "Entry and Profits in an Aging Economy: The Role of Consumer Inertia," Working paper.

Burdett, Kenneth and Kenneth L Judd (1983) "Equilibrium Price Dispersion," *Econometrica*, 51 (4), 955–69.

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

Gentzkow, Matthew, 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.

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.

Kaplan, Greg and Guido Menzio (2015) "The Morphology of Price Dispersion," *International Economic Review*, 56 (4), 1165–1206.

Menzio, Guido (2024) "Markups: A Search-Theoretic Perspective," NBER Working Paper 32888, National Bureau of Economic Research.

Michelacci, Claudio, Luigi Paciello, and Andrea Pozzi (2021) "The Extensive Margin of Aggregate Consumption Demand," *The Review of Economic Studies*, 89 (2), 909–947.

Mongey, Simon and Michael E. Waugh (2025) "Pricing Inequality," NBER Working Paper w33399, National Bureau of Economic Research.

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," Working paper.

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

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

Sangani, Kunal (2024) "Markups across the Income Distribution: Measurement and implications," Working paper.

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

1.A. 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**
Homogeneous Preferences	Uniform Probability	0.1 0.1	0.1 0.1	0.1 0.1
	Non-Uniform Probability	$\frac{0}{18}$ $\frac{0}{18}$	$\frac{1}{18}$ $\frac{1}{18}$	$\frac{2}{18}$ $\frac{2}{18}$
Heterogeneous Preferences	Perfect Separation	0.2 0.0	0.2 0.0	0.2 0.0
	Partial Separation	$\frac{0}{18}$ $\frac{9}{18}$	$\frac{1}{18}$ $\frac{8}{18}$	$\frac{2}{18}$ $\frac{7}{18}$
Preference Type	**Probabilities**	**Good 4**	**Good 5**	**Good 6**
Homogeneous Preferences	Uniform Probability	0.1 0.1	0.1 0.1	0.1 0.1
	Non-Uniform Probability	$\frac{3}{18}$ $\frac{3}{18}$	$\frac{4}{18}$ $\frac{4}{18}$	$\frac{5}{18}$ $\frac{5}{18}$
Heterogeneous Preferences	Perfect Separation	0.2 0.0	0.2 0.0	0.0 0.2
	Partial Separation	$\frac{3}{18}$ $\frac{6}{18}$	$\frac{4}{18}$ $\frac{5}{18}$	$\frac{5}{18}$ $\frac{4}{18}$
Preference Type	**Probabilities**	**Good 7**	**Good 8**	**Good 9**
Homogeneous Preferences	Uniform Probability	0.1 0.1	0.1 0.1	0.1 0.1
	Non-Uniform Probability	$\frac{6}{18}$ $\frac{6}{18}$	$\frac{7}{18}$ $\frac{7}{18}$	$\frac{8}{18}$ $\frac{8}{18}$
Heterogeneous Preferences	Perfect Separation	0.0 0.2	0.0 0.2	0.0 0.2
	Partial Separation	$\frac{6}{18}$ $\frac{3}{18}$	$\frac{7}{18}$ $\frac{2}{18}$	$\frac{8}{18}$ $\frac{1}{18}$
Preference Type	**Probabilities**	**Good 10**		
Homogeneous Preferences	Uniform Probability	0.1 0.1		
	Non-Uniform Probability	$\frac{9}{18}$ $\frac{9}{18}$		
Heterogeneous Preferences	Perfect Separation	0.0 0.2		
	Partial Separation	$\frac{9}{18}$ $\frac{0}{18}$		

Table 1.1: 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 second 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. 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 1.2: Estimated polarization for the different simulations

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.

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. While the polarization estimates at the barcode level are ranging around 0.58, the plots show that polarization is much lower for higher levels of aggregation.

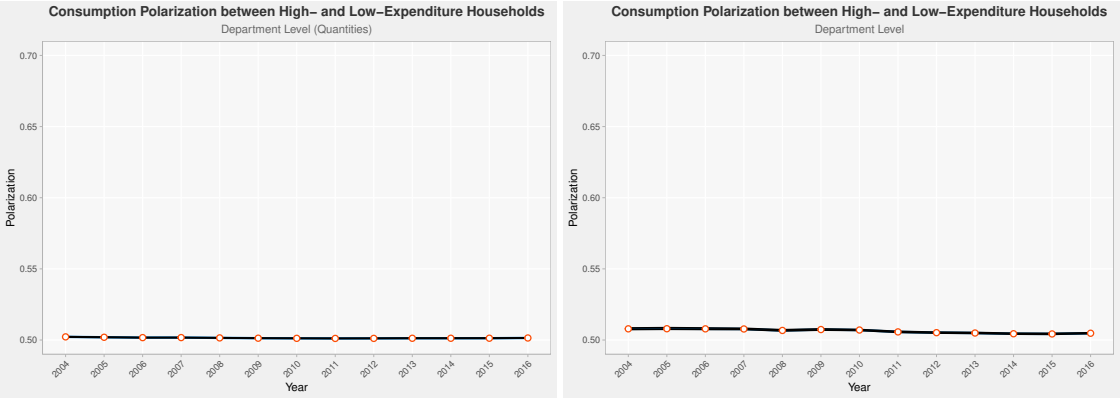


Figure 1.6: Both plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product department level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

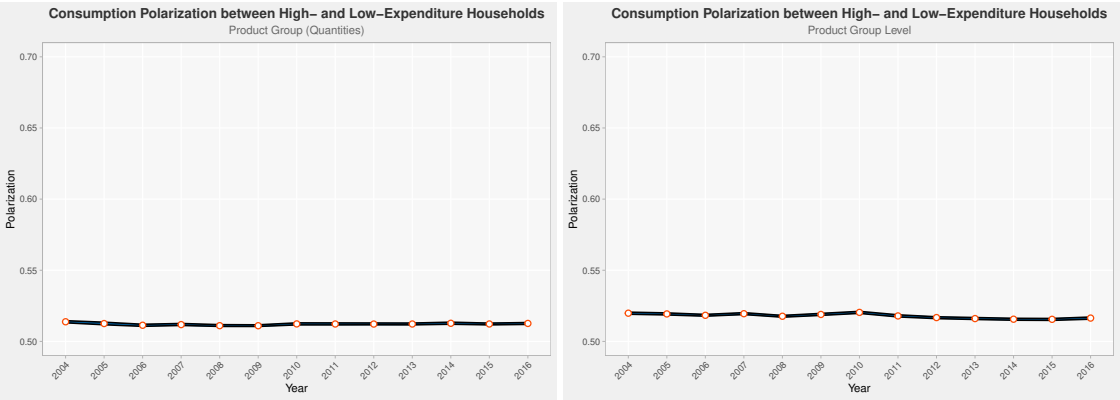


Figure 1.7: Both plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product group level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

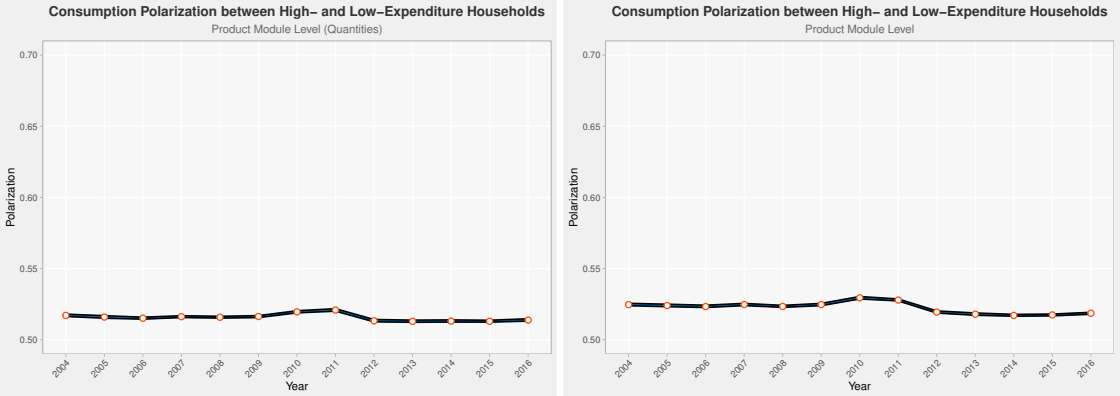


Figure 1.8: Both plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product module level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

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.

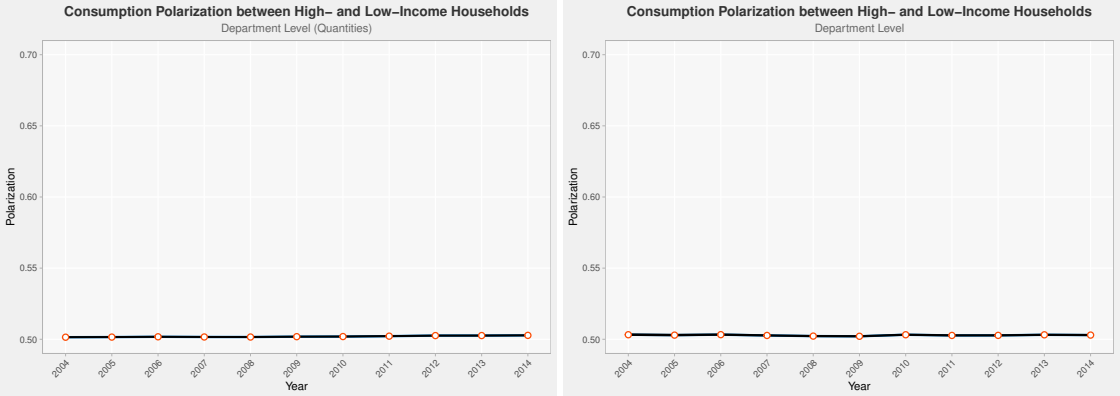


Figure 1.9: Both plots display polarization estimates between the top 20% and bottom 20% of the income distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product department level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

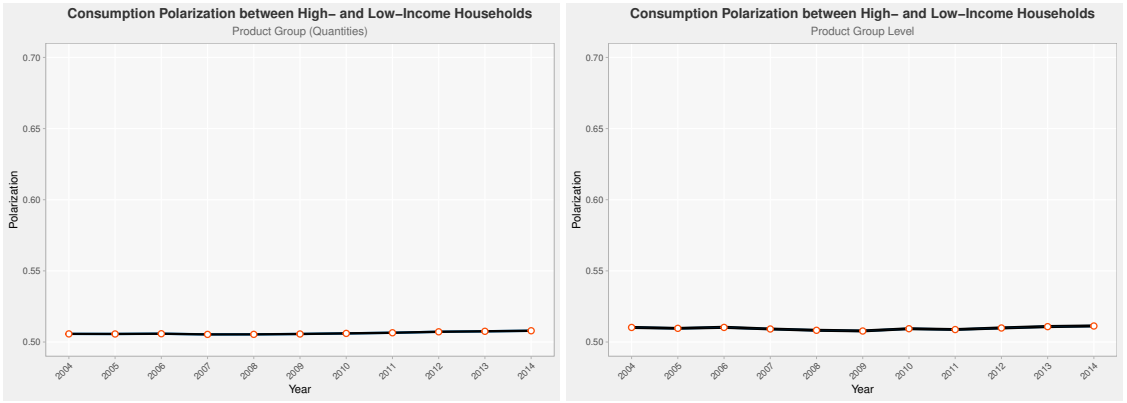


Figure 1.10: Both plots display polarization estimates between the top 20% and bottom 20% of the income distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product group level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

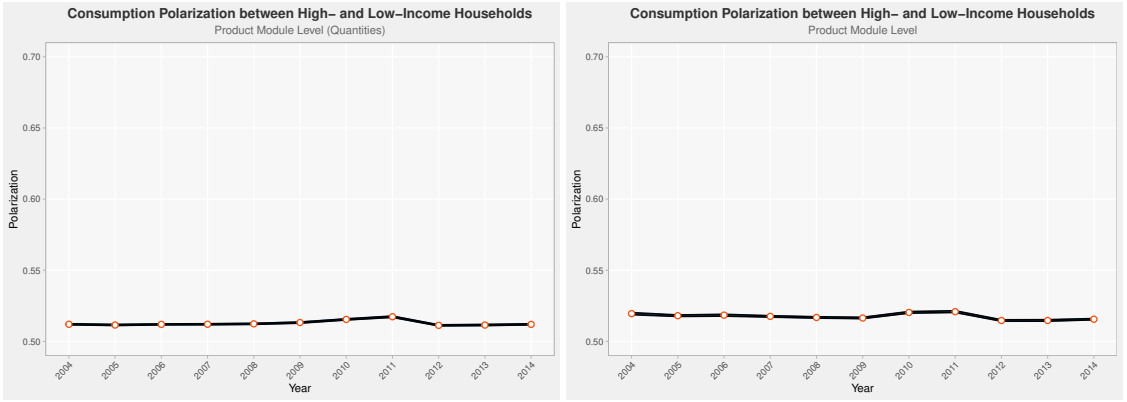


Figure 1.11: Both plots display polarization estimates between the top 20% and bottom 20% of the income distribution. Confidence intervals are 95% and computed as suggested in Gentzkow, Shapiro and Taddy (2019b). Since the confidence bands are relatively tight they are not visible in the plots. All products are defined at the product module level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

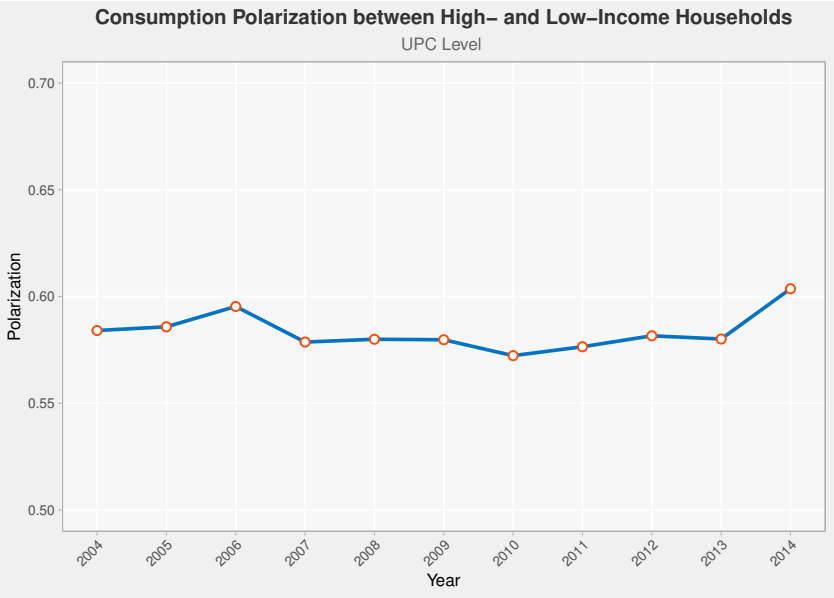


Figure 1.12: The plot displays polarization estimates between the top 20% and bottom 20% of the income distribution. The estimates are based on the within department estimates. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, 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 show 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.

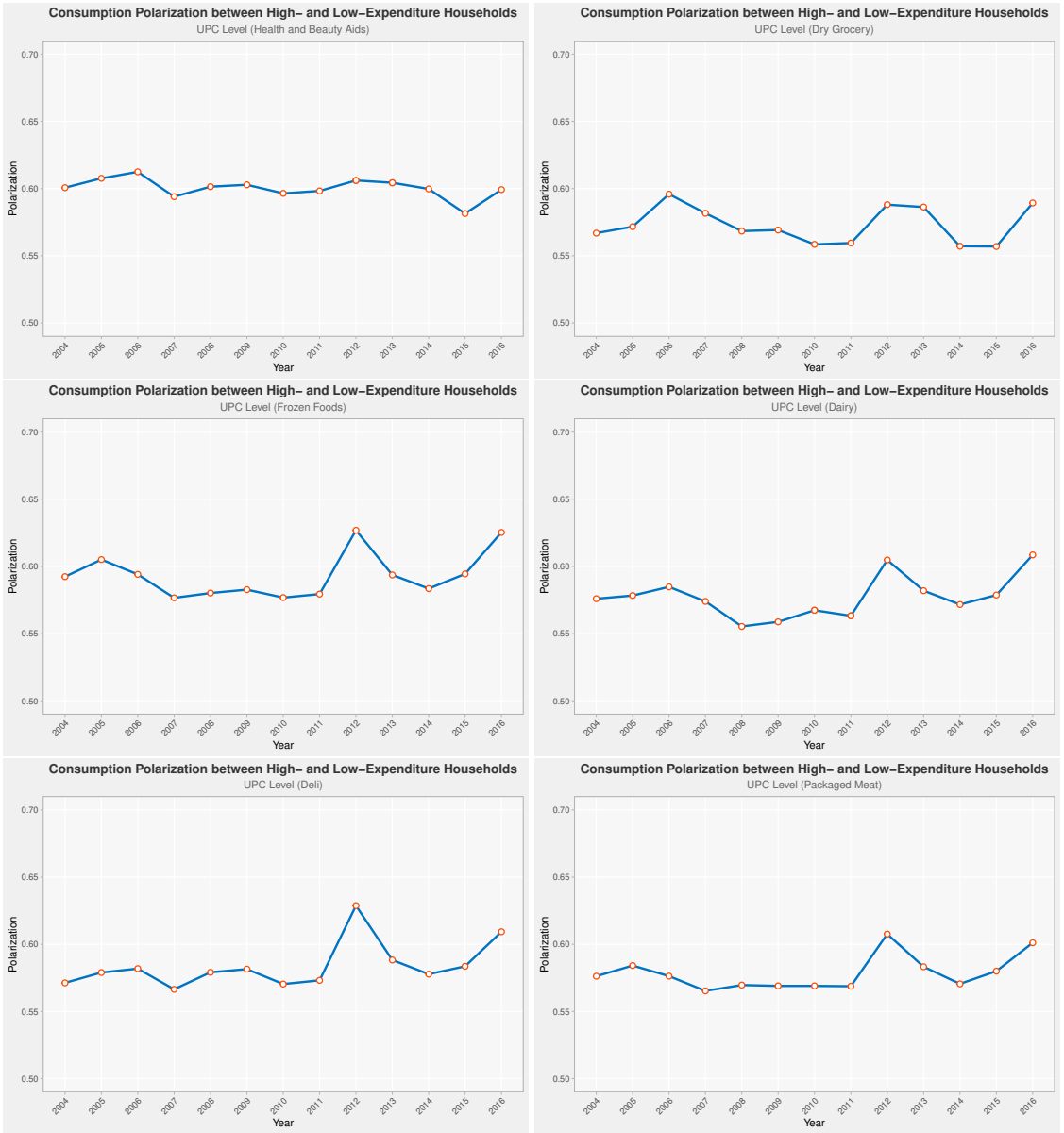


Figure 1.13: The plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

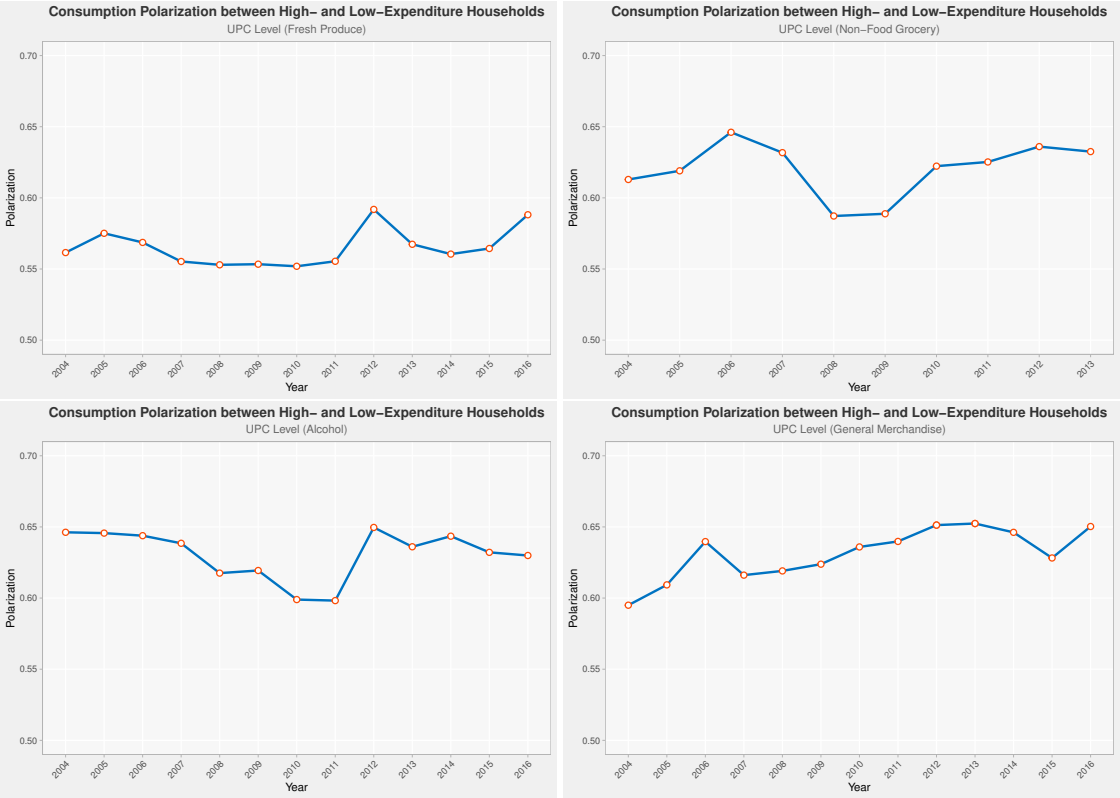


Figure 1.14: The plots display polarization estimates between the top 20% and bottom 20% of the consumption expenditure distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

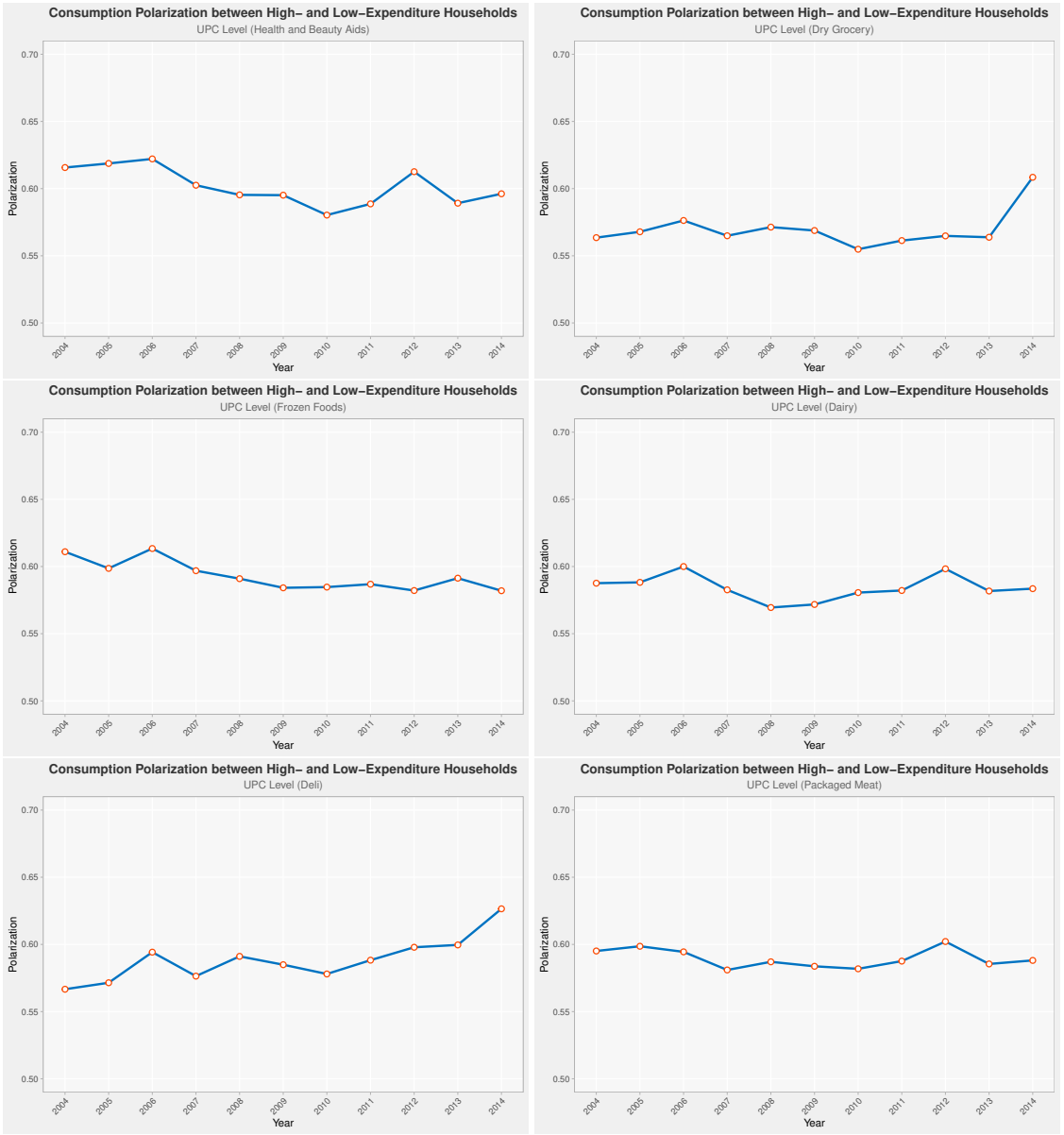


Figure 1.15: The plots display polarization estimates between the top 20% and bottom 20% of the income distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

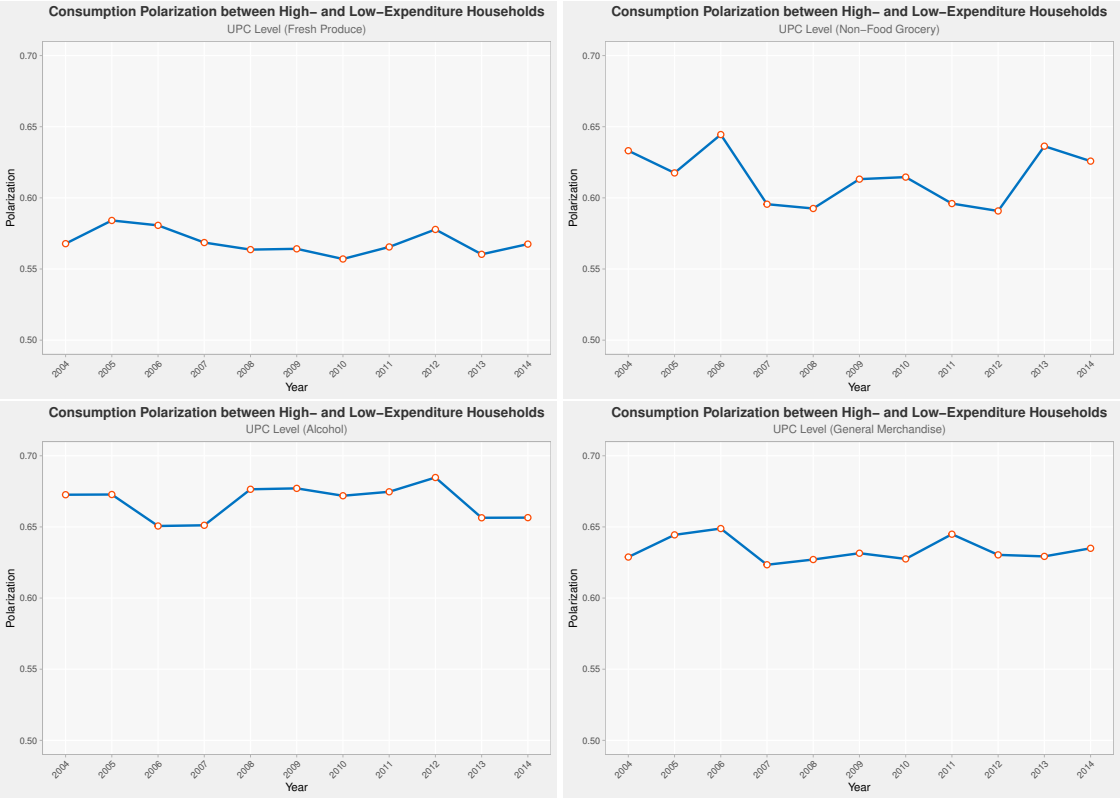


Figure 1.16: The plots display polarization estimates between the top 20% and bottom 20% of the income distribution. No confidence intervals are provided. All products are defined at the UPC level. Polarization estimates can range from 0.5, no polarization, to 1, perfect polarization.

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. Figure 1.17 shows that persistence estimates are not driven by product exit. The figure 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.

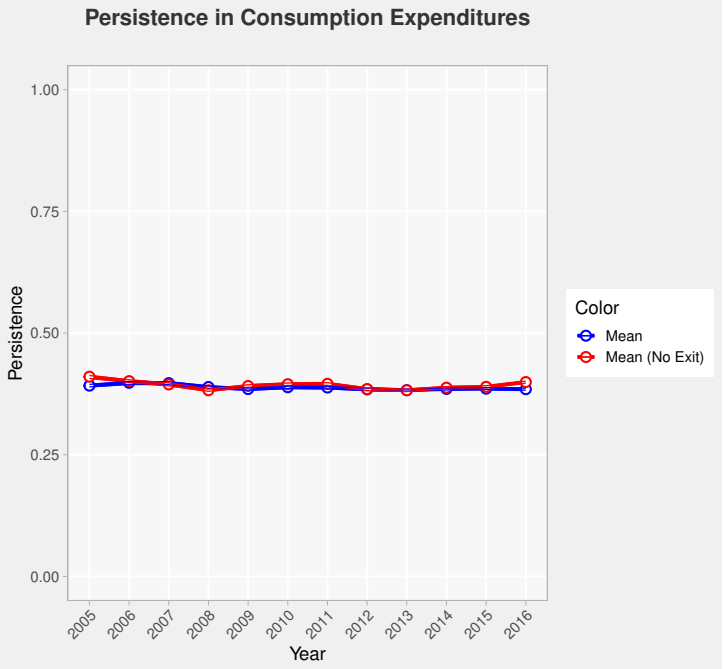


Figure 1.17: Persistence estimates. The figure displays the estimates for the expenditure-based measure. No Exit refers to the estimates where all products are excluded that are not available to buy somewhere in the US in both years considered. All confidence bands are 95% and not visible since they are very narrow.

Figure 1.18 shows the baseline polarization estimates for two groups of households. One group are all households that experience a change in income bracket and the other group 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 other factors than income changes. Changing perspectives, this shows that even in the absence of income changes, consumption baskets change significantly from year to year.

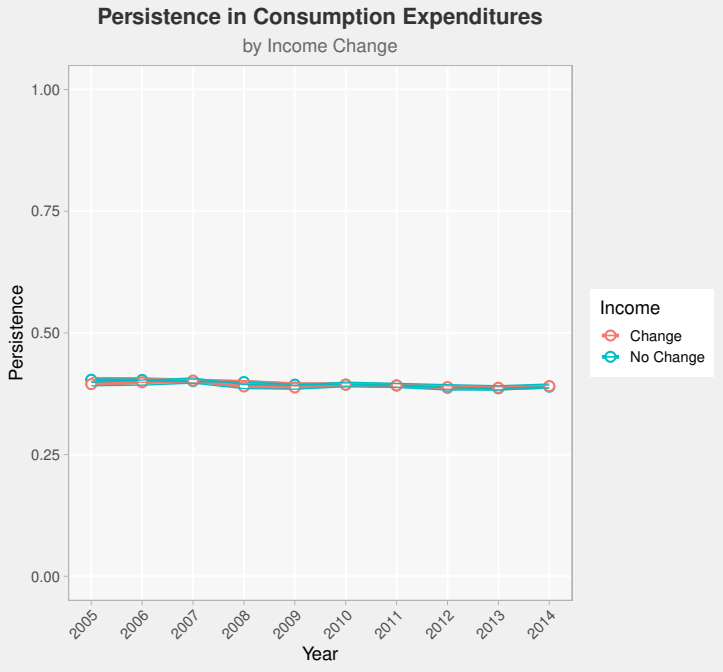


Figure 1.18: Persistence estimates for households that experience a change in income and those that do not. All confidence bands are 95% and not visible since they are very narrow.

Figure 1.19 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 idea is to give a higher weight to households that consume more to get a better feeling 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.

Finally, Figure 1.20 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 would converge to its mean value

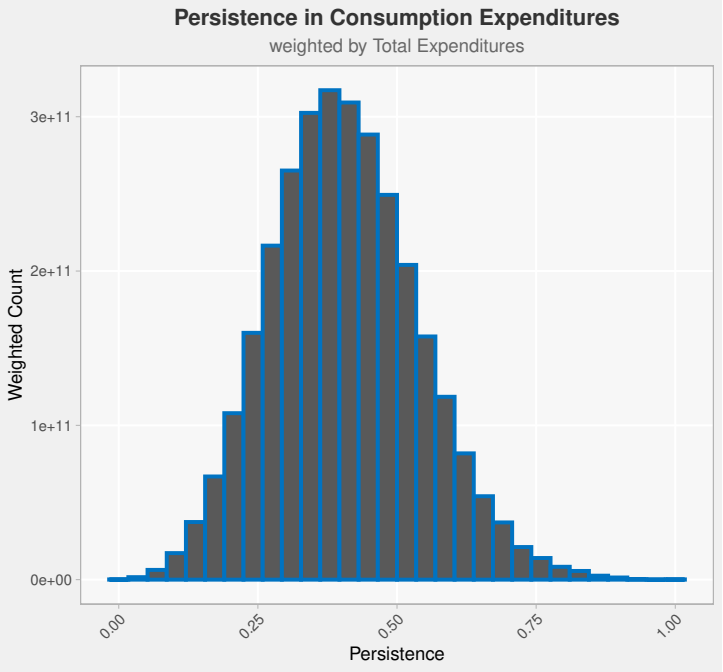


Figure 1.19: Histograms of persistence estimates. The figure shows a histogram of the persistence estimates for the expenditure-based measure weighted by the projection factor as well as the household expenditures.

as the number of consumed UPCs tends to infinity. Put otherwise, 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.

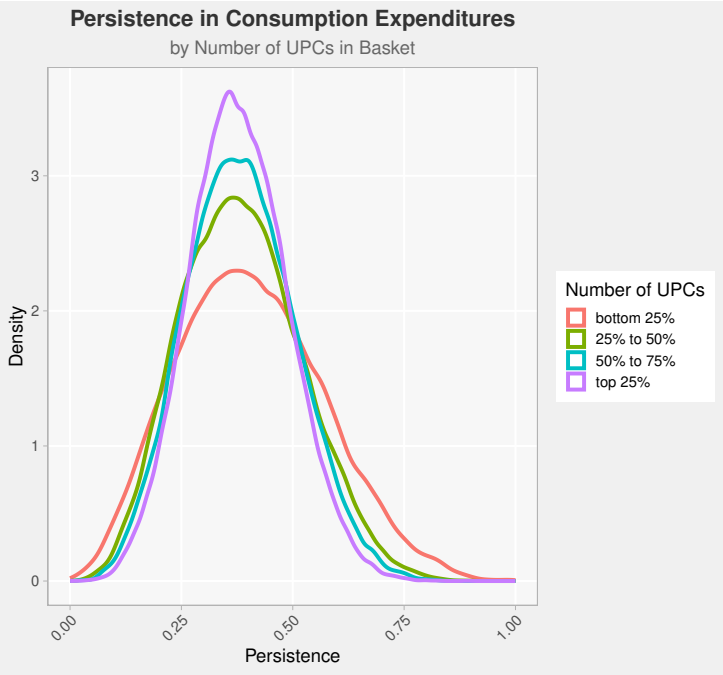


Figure 1.20: Densities of persistence estimates. The figure shows densities of the persistence estimates for the expenditure-based measure, where the households are split into 4 groups according to the number of UPCs in their consumption basket.

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.

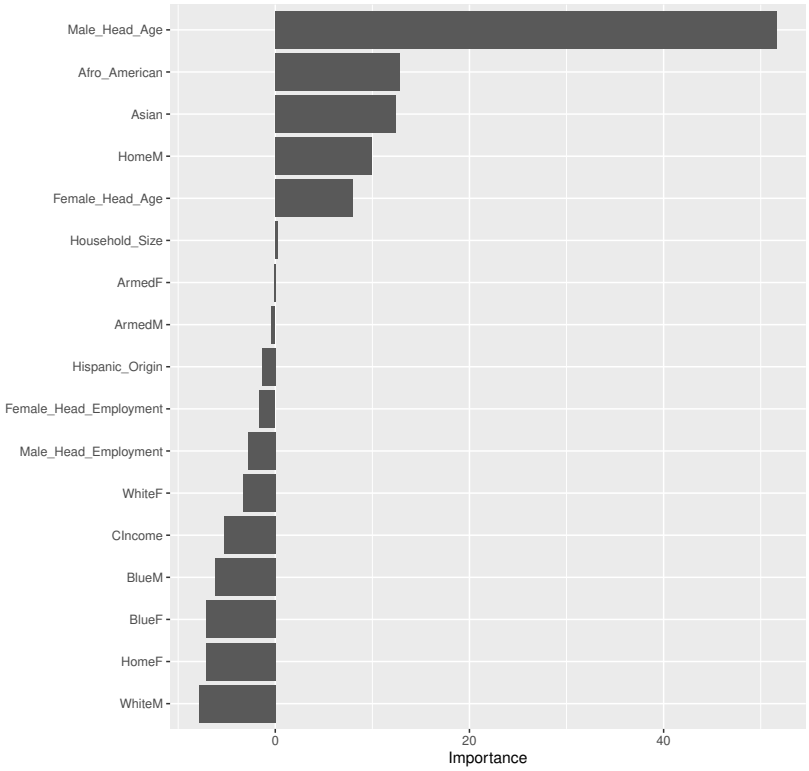


Figure 1.21: Variable importance plot for the random forest model

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. When examining the size of

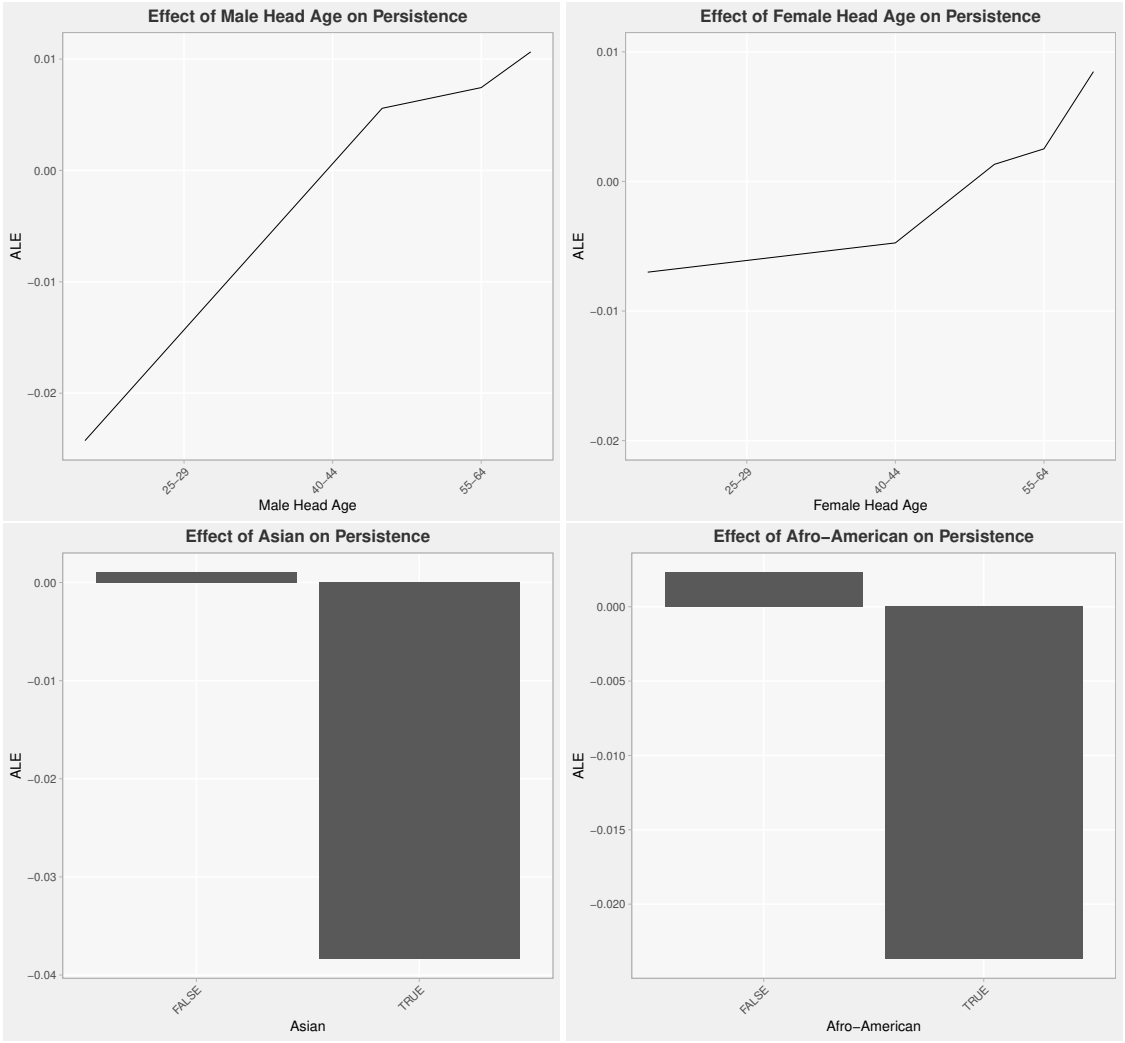


Figure 1.22: Accumulated local effects estimates for the random forest model

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

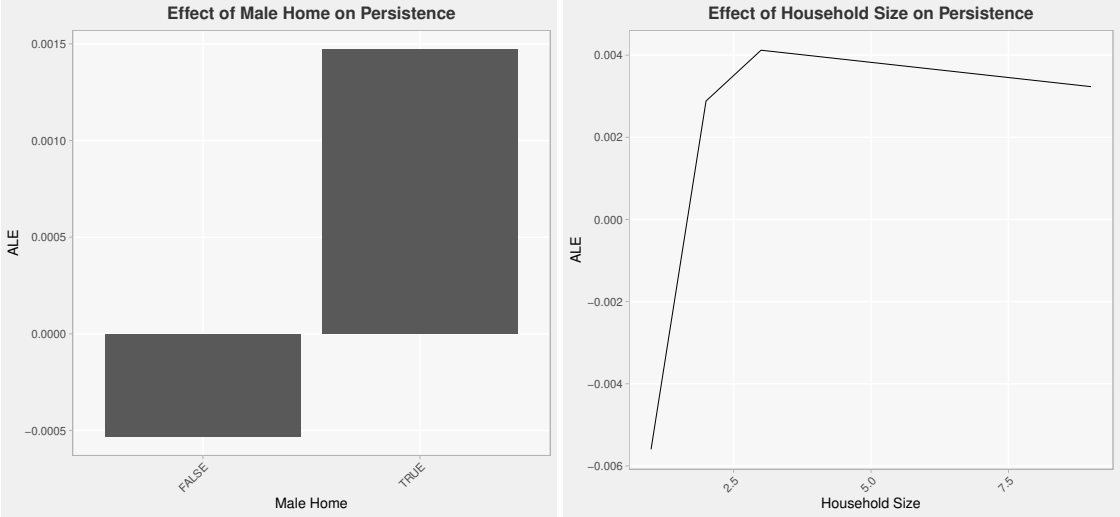


Figure 1.23: Accumulated local effects estimates for the random forest model

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.

Chapter 2

Puzzle or no Puzzle? - Reexamining effects of monetary policy shocks on prices using consumer-level price data

This paper reexamines the impact of monetary policy shocks on prices, acknowledging that paid and posted prices are not the same. I use consumer microdata from the Kilts-NielsenIQ Consumer Panel and shocks identified by Jarociński and Karadi (2020) to analyze the effects of monetary policy. I find that an undecomposed contractionary shock lowers paid prices, but a pure monetary policy shock unexpectedly increases them. Households adjust their price search behavior differently depending on their income level, with low-income households increasing their search effort relative to the high-income group. While there is no evidence of the overall product quality of households' consumption baskets changing in response to a shock, within-department analysis reveals that quality adjustments take place. The study highlights the importance of considering paid prices and household heterogeneity in monetary policy analysis.

I. Introduction

At the heart of monetary policy models lies the idea that the central bank can, by conducting conventional monetary policy, affect the level of prices within the economy. While the effect of monetary policy shocks on inflation rates and, more generally, on prices has been extensively studied, the results obtained from these studies have varied greatly. Where some studies document the expected effect—namely, that a shock that increases the interest rate leads to lower inflation—others show the existence of a so-called price puzzle, documenting no effect or even a reverse effect.

One feature that is common to all of these studies is that they rely on data about prices posted by sellers, neglecting the fact that these prices are not identical to those paid by customers. More importantly, paid, not posted, prices are the relevant measure for economic activity as well as household welfare and therefore should be the focus for policymakers and researchers alike. Paid prices determine household expenditures and thereby potential, as well as actual, consumption. Similarly, it is also paid, not posted, prices, along with the quantities sold at different price points, that ultimately determine retailer revenue and profit.

In the consumption price search literature, there has been significant effort demonstrating how different the prices paid by individual customers actually are. Even when controlling for identical products down to the barcode level, the prices paid can vary greatly, as shown by Aguiar and Hurst (2007), Kaplan and Menzio (2015), and Pytka (2024). More importantly, these differences are not random but are highly correlated with observable socioeconomic characteristics of households. Differences between paid and posted prices arise once one considers that customers might have more than one price offer for a given product. This occurs because customers will spend effort to find better prices and then buy the products at the best price found, while sellers will trade off higher per-unit profits against a higher volume of sales. In turn, this will lead to more trading activity at lower offered prices. This mechanism lies at the heart of the seminal price search model developed by Burdett and Judd (1983). Besides being caused by search frictions, dispersion in prices, especially at higher levels of aggregation, can also be the result of price-setting frictions, differences in product quality, or similar concepts.

Genuine differences in product quality between varieties of similar products, with higher-quality varieties being more expensive, will lead to differences in prices when these varieties are viewed jointly.

The aim of this project is to incorporate these findings into the analysis of monetary policy shocks. I use barcode-level consumer microdata from the Kilts-NielsenIQ Consumer Panel, along with established monetary policy shocks from Jarociński and Karadi (2020), to demonstrate how three key variables of interest respond to the shocks. These are paid prices, relative prices, and the average quality of goods in households' consumption baskets. All three of these variables are crucial to get a full picture of how households respond to the shocks, as well as how and why paid prices change. Paid prices show if there is a direct effect of the monetary policy shocks on the prices paid by households. Crucially, a change in the average paid price can be caused either by retailers adjusting the prices charged for a product, by households adjusting their search effort, or, finally, by households changing the composition of their consumption basket. Looking at the responses of these variables jointly allows policymakers to get a complete picture of how households respond to a monetary policy shock.

Analyzing how relative prices change in response to the monetary policy shock will allow conclusions to be drawn about changes in households' price search behavior in response to the shock. This is the case because relative prices reflect how expensive a specific product is for a household compared to the average price of this good within the whole market. A change in this relative price will then reflect a change in the relative price search effort of a type of household compared to other households within the economy. The response in the quality of households' consumption baskets will instead allow identification of whether the composition of the consumption baskets changes in response to the shocks. Additionally, the direction of the response will indicate whether households are substituting products with higher or lower quality products.

I start by establishing the effect on the average price paid. The average price responds to a contractionary shock with a decrease of 0.12% on impact, and the response is stronger for high-income households. When weighting products by their expenditure share, the response is less pronounced. Additionally, there is a significant difference between income groups. While the price response is deflationary

for the high-income group, the response for the low-income group is inflationary. For relative prices, I find that a contractionary shock leads to high-income households paying relatively higher prices for the same goods as low-income households. Finally, I find evidence of product substitution in response to a contractionary shock.

Literature Review. This project primarily contributes to the literature on monetary policy. It is most closely related to the empirical studies of monetary policy shocks. There has been a vast amount of work done using different identification schemes and data sources to identify the effects of monetary policy shocks on macroeconomic aggregates. While identification has been based on narrative approaches (Romer and Romer, 2004) as well as different forms of zero restrictions combined with Cholesky decompositions (Christiano et al., 1999), the current state-of-the-art is to use fluctuations in high-frequency financial data in a short window around the announcement of new monetary policies (Gertler and Karadi, 2015) to identify the shocks.

The empirical evidence produced by these studies has been mixed. While some of the studies have been able to show the expected effect—that a contractionary monetary policy shock leads to lower inflation—other studies have found a price puzzle, showing no effect or that a contractionary shock actually increases inflation. Ramey (2016) gives an overview of the most important studies and also highlights which studies find the expected effect and which document the presence of a price puzzle. A more detailed discussion on why within this study I work with the shocks of Jarociński and Karadi (2020), which are constructed similarly to Gertler and Karadi (2015), is provided in the data section later on.

In addition, this project is related to the literature on price dispersion and price search. For instance, Aguiar and Hurst (2007) develop an index to capture how prices paid for identical products differ between different households. They show that price differences are non-random but connected to households socio-economic background. Kaplan and Menzio (2015) describe dispersion within prices in more detail and Kaplan and Schulhofer-Wohl (2017) show that there is also dispersion within inflation indices between different households'. More recently Pytka (2024) showed that higher earning households pay higher prices for identical goods.

II. Empirical Analysis

One point that is common to most if not all studies of monetary policy shocks is that they primarily use inflation data directly from the statistical offices or a mixture of price data from the statistical offices and quantities for example from the survey of consumer expenditures. These offices usually collect price data from stores or websites either monthly or bimonthly, and the data is then aggregated for similar items. This has two major drawbacks. First, by collecting price data only from the seller side, the fact that posted and paid prices are different is ignored. Second, by collecting prices and quantities separately, the fact that there is huge dispersion in prices and consumed items between households is neglected.

The reasons for the emergence of a difference between paid and posted prices as well as multiple prices for identical goods within a market are perfectly illustrated by the model in Burdett and Judd (1983). The main idea is based on the observation that households will often not buy a product at the first observed price but will search for a better deal. This opens the possibility for the sellers to trade off higher sales volume against higher per unit profits. At the highest price posted by the sellers, only those households that find no other price offer will purchase the product from this seller. As the price decreases, the probability of finding a better price for the product decreases and therefore the volume of sales increases. Hence, different numbers of trades occur at different prices, leading to a distribution of paid prices that differs from the posted price distribution. Importantly for the sellers, there is no incentive to change prices as profits are equalized between all posted prices due to the fact that higher sales compensate for lower per unit profits. Overall, one can think of the posted price distribution as arising from the retailers' pricing decisions, while the paid price distribution is the result of the interaction between the posted price distribution and the households' price-search behavior.

More crucially, using paid instead of posted prices might change the results when analyzing the effect of monetary policy shocks. As posted prices primarily reflect changes within the retailers' pricing decisions, one misses that there might be changes in household behavior that do not directly translate into posted price changes while still affecting the prices paid by households. For instance, some

households might adjust their price-search behavior in a way that compensates or even overcompensates for the effect that monetary policy has on posted prices. In addition, some households might respond instead by changing the composition of their consumption baskets, switching either to a different quality level of a consumed product or changing the product types consumed. Also note that while responses within posted prices might be delayed by price setting frictions, households are free to adjust their search behavior instantly.

Not only are paid prices the more appropriate measure to capture to the full extent the effect of monetary policy shocks, they are also intrinsically the more relevant metric. As economists, we are ultimately interested in the effects of policies on household welfare. It is paid prices, not posted prices, that determine the amount of consumption of a household. While one might argue that lower paid prices may indicate that a more intensive price search has taken place (which has a utility cost), assuming rational consumers implies that the overall welfare effect of the search remains positive. Therefore, paid prices can be seen as an approximate measure of household welfare. Additionally, paid prices are a more relevant metric for profits and markups as those depend not only on the prices posted but also on the amount of trade happening at those prices.

If one takes the view of a policymaker or central banker, one comes to a similar conclusion. In addition to the fact that paid prices determine household welfare, they are relevant for the expectations formed by households about future prices. Therefore, insofar as a policymaker is concerned with the effect monetary policy shocks have on household price expectations, they should care much more about the response of paid prices than posted prices.

Furthermore, paid prices cannot be approximated by posted prices to a satisfactory degree. Neither is the distribution of posted prices a sufficient statistic for the distribution of paid prices, nor are changes in paid prices a sufficient statistic for changes in posted prices. As long as the researcher or statistician does not have exact knowledge about households' search and buying behavior, one cannot judge how changes in posted prices will translate into changes in paid prices. This is because households have multiple interconnected ways to respond to those changes, which can potentially affect paid prices in different directions.

Besides the issue of using posted instead of paid prices, consumption baskets

used to study the effect of monetary policy shocks are typically constructed by assigning one price per item and fixing a representative basket over a longer period. From the literature on price dispersion, it becomes clear that this approach has two main issues. First, even when conditioning on goods being identical, there is significant dispersion in prices paid by different households, as shown for example by Kaplan and Menzio (2015). For example, higher-income households tend to pay, on average, higher prices for the same goods compared to lower-income households. In addition to dispersion in prices, there is significant dispersion in household-level inflation rates as shown by Kaplan and Schulhofer-Wohl (2017). As D'Acunto et al. (2021) have demonstrated, this dispersion in realized household-level inflation rates then translates into dispersion in inflation expectations, which in turn influence household behavior. Second, as demonstrated by Pytka ⓡ Runge (2025), household-level consumption baskets are highly unstable at the barcode level. Taken together with the fact that there are systematic price differences between households, this suggest that price or inflation indices constructed from posted price data obtained from the statistical offices will severely mismeasure actual household inflation at a granular level, even if average inflation is captured correctly.

This study is designed to address these weaknesses. By using detailed consumer microdata, I am able to account for the different prices paid by different consumers, as well as for their varying consumption habits and preferences. I utilize the identified monetary policy shocks from Jarociński and Karadi (2020), which offer the most precise method to isolate the effect of monetary policy surprises.

I also show in Runge (2025) that quality differences between products can be thought of as a combination of genuine quality differences and additional spurious differences generated through the interaction of search frictions and data sparsity. Nonetheless, I will document how product quality responds to shocks in monetary policy and, assuming that the spurious part of the differences is independent of the shocks, the results should not be affected by this issue. Even though Pytka ⓡ Runge (2025) show evidence that there is at most a relatively weak common product quality ladder for all households, I will still assess how quality responds to the different shocks as well as whether there are differences in the response between the top and bottom 20% of the income distribution. This question is

still relevant as Pytka ⓡ Runge (2025) are only concerned with unconditional differences in consumption baskets, while in this study the difference in response to an exogenous shock is analyzed. In addition, I look at both the overall quality of the basket as well as quality within individual departments.

A. Data Description

For this study, I use primarily data from two different sources. The monetary policy shocks are taken from Jarociński and Karadi (2020) and the consumer microdata is taken from the Kilts-NielsenIQ Consumer Panel (KNCP).

The KNCP meticulously records grocery purchase details from an evolving panel of American households, ranging from approximately 40,000 in the years 2004-2006 to 60,000 from 2007 onward. By utilizing either in-home scanning devices or mobile applications, panel participants provide NielsenIQ with comprehensive purchase records from various retail outlets nationwide, linking each product purchase to a distinct shopping event. Participants also submit socio-demographic data annually, with NielsenIQ providing household weights to align the sample with broader U.S. economic demographics. The dataset encompasses 54 distinct geographic markets, identified as Scantrack markets, incorporating all available data spanning 2004 to 2014. Throughout this timeframe, the KNCP amassed data on 630 million transactions involving close to 2 million unique products, identified by their barcodes (UPCs), and assembled from 87 million documented shopping excursions.

Jarociński and Karadi (2020) use a high-frequency identification scheme to extract monetary policy shocks from financial data collected in a short time window around the announcement of new monetary policy. The main idea is that if the window is short enough, then the only change observed is caused by the announcement of the new policy.

In addition to the shocks directly derived from the financial market changes, they provide a decomposition into two components: one called the pure monetary policy shock and one called the central bank information shock. The idea behind this decomposition is that the raw shock contains two components. On the one hand, there is what is usually understood as a monetary policy shock, and on the

other hand, there is some information that was known to the central bank but not to market participants. The first part refers to monetary policy in a narrow sense, while the second part is simply the revelation of additional information to the market and is not encapsulated in the usual understanding of a monetary policy shock.

The shocks are constructed as follows. In the first step, surprises in financial market instruments are identified. This is done by collecting the changes in the relevant financial market instruments within a window around the monetary policy announcement. If there is more than one announcement within a given month, the surprises are aggregated. These surprises together with a set of macroeconomic outcomes are then used as the basis to identify the structural shocks. The structural shocks are identified using two main assumptions. First, only the structural shocks are allowed to affect the surprises in the financial market instruments. The second assumption differentiates between the two structural shocks by restricting the way they are allowed to affect the surprises and the overall stock market. While both shocks lead to an increase in surprises, the monetary policy shock affects stock market prices negatively, while the central bank information shock affects them positively. The effect on other macroeconomic outcomes is left unrestricted. In the following, I will refer to the change within the financial market instruments as the monetary policy shock or undecomposed shock, while the two structural shocks are referred to as pure monetary policy shock and central bank information shock.

B. Estimation Methodology

Before discussing the results of my estimation, I will first present how the variables for which I want to estimate impulse responses are defined and then describe how the responses are estimated. The variables I want to investigate are average paid prices, relative prices, inflation as well as product quality.

For each of the products within the dataset, I compute the average paid price. Denote by $p_{i,j,t}$ the price of good j purchased at time t by household i, and $q_{i,j,t}$ as the quantity of good j purchased at time t by household i. Furthermore let $\psi_{i,t}$ be

the projection-factor of household i at time t. Then the average price is given by:

$$\bar{p}_{j,t} = \frac{\sum_i p_{i,j,t} q_{i,j,t} \psi_{i,t}}{\sum_i q_{i,j,t} \psi_{i,t}}$$

I measure relative prices using the index proposed by Aguiar and Hurst (2007). This index has the benefit of being simple to compute as well as having a clear and intuitive interpretation. The index is constructed by comparing the expenditures of a household to the hypothetical expenditures it would have if all goods were bought at average prices. Hence, a value below 1 indicates that the household payed on average relatively low prices, and a value above 1 indicates that the household payed on average relatively high prices. To define the index more formally, let $Q_{i,t}$ represent the set of UPCs consumed by household i at time t. Given this, the index is defined by:

$$p^r_{j,t} = \frac{\sum_{j \in Q_{i,t}} p_{i,j,t} q_{i,j,t}}{\sum_{j \in Q_{i,t}} \bar{p}_{j,t} q_{i,j,t}}$$

When interpreting the relative price index as well as impulse responses later on, it is crucial to keep in mind that the index only reflects the prices paid by a household relative to the prices paid by other households, and not the absolute level of prices. This means that if, for instance, all paid prices double while the consumed quantities stay the same or change by the same proportion for all households, the relative price index would remain unchanged.

Inflation is measured following the index used in Kaplan and Schulhofer-Wohl (2017). The index is constructed at a quarterly level. Formally, it is defined by:

$$\pi_{i,t+4} = \frac{\sum_{j \in \{j: q_{i,j,t} > 0, q_{i,j,t+4} > 0\}} p_{i,j,t+4} q_{i,j,t}}{\sum_{j \in \{j: q_{i,j,t} > 0, q_{i,j,t+4} > 0\}} p_{i,j,t} q_{i,j,t}}$$

One important feature of this index is that in contrast to the official inflation measure, the consumption basket changes from period to period. This has the advantage of allowing the use of the maximum amount of data to compute inflation for each household and each period without having to impute prices. Note that the comparison is always between prices for the same quarter one year ago, so there is

no concern about seasonal changes within the basket affecting inflation or that the inflation measure picks up seasonal variations in prices. In addition, by allowing the consumption basket to change over time, it incorporates possible substitutions undertaken by the households to compensate for different rates of price increases over time. If for example households substitute away from products with high rates of inflation to products with lower inflation rates, an index with fixed consumption baskets would severely overestimate inflation at the household level. Also as Pytka ⓡ Runge (2025) show, household consumption baskets when products are defined at the barcode level are highly unstable. Therefore, a fixed basket together with the requirement that the household buys the goods in both periods would lead to a small subset, that might be biased, being incorporated into the computation of the final inflation index. This problem is mitigated by the fact that the basket is defined at the household level and changes each period. Nonetheless, the preferred measure to assess price changes is the average paid price of a product, which is computed at the product level.

To examine product quality, this paper draws on the approach of Argente and Lee (2021), which is based on the assumption that the average price of a product reflects the quality of a product. The idea is to measure quality by comparing the average price of a product to the average prices of similar products. If the product is expensive compared to similar products, it is of higher quality, because people are on average willing to pay a higher price for it. To formally define the product quality measure, denote by $p_{j,t}$ the price of product j at time t and by $\hat{p}_{m,t}$ the average price of product module m at time t, where m is the product module to which product j belongs. The quality of product j is then defined as:

$$R_{j,t} = log(p_{j,t}) - log(\hat{p}_{m,t})$$

Given the assumption that the average price of a product captures its quality, this quality measure captures whether a given product is above or below the average price within its product group and to what extent and therefore where it is located within a product quality ladder. By using the average price within module as a normalization, the quality measure becomes comparable across modules. To arrive at a proxy for the overall quality of a given household i's basket, define $\psi_{i,j,t}$ as

the expenditure weight of household i on product j at time t. Then:

$$Q_{i,t} = \sum_j R_{j,t}\psi_{i,j,t}$$

Impulse responses are estimated by employing local projections. To be precise, all impulse responses are estimated using weighted panel local projections with household-level fixed effects:

$$log(y_{i,t+h}) - log(y_{i,t-1}) = \alpha_i + \beta shock_t + \gamma x_{i,t}$$

where y is the measure of interest, α_i is a fixed effect, β represents the effect of the shock, and $x_{i,t}$ is a vector of controls. All confidence intervals are 95% bands. To measure monetary policy shocks, I use the identified shocks provided by Jarociński and Karadi (2020). For the estimation, I use both the undecomposed change in the interest rate swaps as well as the decomposition into a pure monetary policy shock and a central bank information shock. Fixed effects are at the household level except for the average price responds where fixed effects are at the UPC level. As controls, I include a vector of economy-level variables allowed to have a contemporaneous effect on inflation, namely industrial production, a house price index, and the unemployment rate. For all measures except relative prices, I estimate responses for all households simultaneously as well as seperate responses for the top and bottom 20% of the income distribution. For relative prices, I only estimate the response seperately for the two income groups. Estimating one response for all households jointly would not be sensible, since relative prices measure the degree to which a household gets goods cheaper or more expensive compared to other households.

C. Results

I start by estimating the response of the average paid price to a monetary policy shock. This will establish a basis for the interpretation of later results. All responses are estimated at a quarterly frequency and up to a duration of seven quarters. I estimate the impulse responses by imposing symmetry and present the results for the contractionary shock value. Additionally, I normalize the results to

show the effect of a one standard deviation shock. In the main body, I present results for the undecomposed shock and relegate a discussion of the decomposed shocks to the appendix, while highlighting major differences in results within the main text.

The analysis will start by establishing the effect of the shocks on the average prices paid by households. The response of the average price can be deconstructed into three parts. First, average paid prices change due to households adjusting their price search behavior. Second, firms change their posted prices and third, households change the composition of their consumption basket. Taken together, these three factors will govern how the average price of a product changes. The first two channels are directly affecting the average price, while the third one is more indirect and works thought changing the composition of demand for the product.

Therefore, after establishing the effect of monetary policy shocks on average prices, I will look mostly look at two further measures to better address each of those channels separately. To get a sense of whether households adjust their search behavior, I will next look at relative prices. Since relative prices are invariant to linear transformations of the posted price distribution, they can be used to identify differences in the adjustment of search behavior between high- and low-income households. Subsequently, I will very briefly comment on how inflation rates respond to the shock. A full description of the inflation rate responds is relegated to the Appendix, since due to the high degree of basket instability inflation rates are less informative then average prices. Finally, to complete the picture, I will look at the response of the quality of products consumed by households, both for the overall basket as well as for each department separately. Changes in the quality of the consumption basket of the household or of the goods consumed from a specific department are a clear indicator of product substitution, while the absence of changes cannot be interpreted as no substitution taking place.

Figures 2.1 and 2.2 show the response of the average paid price in general as well as split by income group, with two different weighting schemes. In one specification, all products are weighted equally, while in the second one, expenditure shares are used as weights. The main features can be summarized as:

Fact 1. (Average Price Response) *A contractionary monetary policy shock leads to a decrease in the average paid price. In particular:*

1. *The average price drops by about 0.12% on impact when weighting products equally and by about 0.03% when weighting by expenditure shares.*

2. *For the specification using expenditure shares for weighting there is a price increase after quarter 5*

3. *The response of high-income households is more deflationary.*

4. *The difference in responses between the high- and low-income group is much stronger when weighting products bei expenditure shares.*

This establishes that indeed not only posted but also paid prices are affected by monetary policy changes. The shape of the response depends on the weighting of the individual products in the estimation procedure. As can be seen from the estimated response, for both specifications there is a small but significant drop in the average paid price. In response to a contractionary one standard deviation shock, prices drop on impact by about 0.12% for equal product weighting and 0.03% for expenditure weighting. For the equally weighted specification, the effect is persistent over all seven quarters and reaches its peak after five quarters, with a drop of about 0.25%. The magnitude of the response is slightly smaller but roughly in line with the effect on inflation estimated in Jarociński and Karadi (2020), who find that, on impact, the price level drops by 5 basis points. When using expenditure weighting, the peak drop is already reached in quarter 1, and the response becomes inflationary after quarter 5. Since the responses are overall more deflationary for the specification with equal weighting, products with higher expenditure shares react with a lower price decrease or a price increase compared to products with lower expenditure shares. In both specifications, the decomposed shocks lead to price increases in response to a contractionary shock, as can be seen in Appendix A.

Estimating responses instead of using average paid prices for inflation rates changes the results significantly. As can be seen in Appendix A, a contractionary shock is then estimated to have an inflationary effect. As there are significant issues

with constructing inflation indices at the household level, as highlighted before, the responses estimated for the average paid price using expenditure weighting are a better indicator of the evolution of prices at the household level in response to a monetary policy shock.

Since paid prices are a combination of the prices offered by retailers and the search effort expended by consumers, one can expect paid prices to vary significantly between relatively rich and relatively poor households. These differences have also been established within the empirical literature. Similarly, one can expect households to respond differently to changes induced by monetary policy depending on their income. To investigate if this is the case, I estimate the response of the average paid price for the top and bottom 20% of the income distribution.

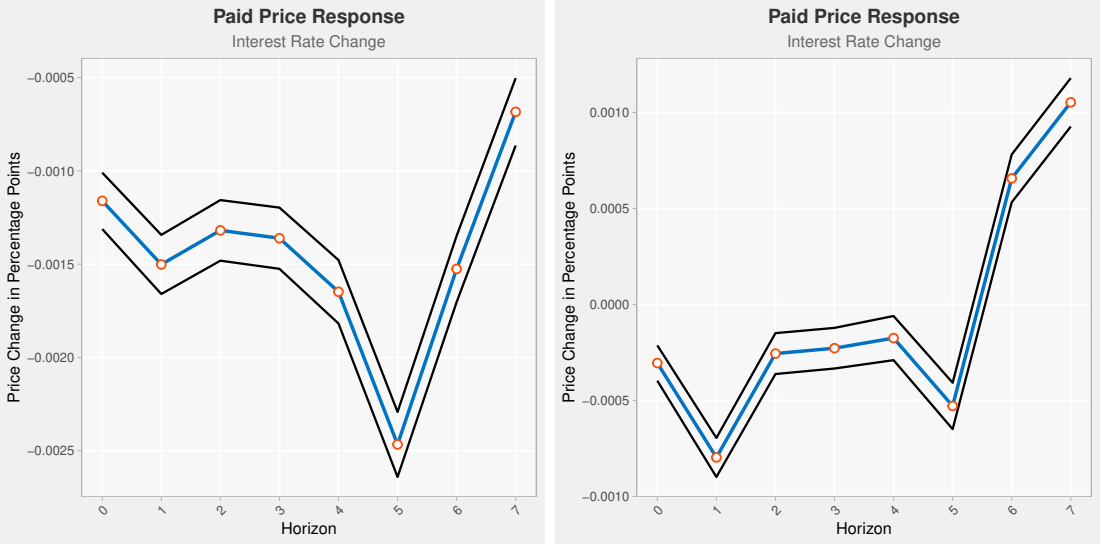


Figure 2.1: Estimated impulse response of the average paid price of a product including 95% confidence bands. For the plot on the left all products are weighted equally in the estimation and on the right they are weighted using expenditure shares. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

As can be seen from the estimated responses, for the equally weighted specification the overall picture is relatively similar to the response estimated for the whole sample. There is a significant negative response of about the same size for both groups. While there is no significant difference between the estimated

responses up to the third quarter, starting in the fourth quarter, there is a clear statistical difference between the estimated responses. The effect of the shock fades out faster for low-income households, and the low point in the price drop is lower for the high-income group. For the specification using expenditure weighting, the case is quite different. Starting from quarter 1 onward, there is a clear difference between the estimated responses for the two groups. While the response of the high-income group is mostly deflationary, the price response for the low-income group is clearly inflationary starting from quarter 1. The difference in responses can be caused either by differences in the consumption baskets between the two income groups or by different paid price responses for the same goods.

Figure 2.2: Estimated impulse response of the average paid price of a product for the top and bottom 20% of the income distribution including 95% confidence bands. For the plot on the left all products are weighted equally in the estimation and on the right they are weighted using expenditure shares. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

To disentangle the two, I will next look at the response in relative prices, which addresses the second point. If there is no significant response in relative prices, then this suggests that the difference originates from differences in the consumption baskets. The difference in the average price responses between the two income groups is then caused by the fact that both types of households consume different

kinds of goods in the first place. A significant relative price response, however, would suggest that at least part of the difference is due to prices paid for the same goods developing differently in response to the shocks for both groups. The relative price response can be summarized as:

Fact 2. (Relative Price Response) *A contractionary monetary policy shock leads to a relative price decrease for low-income households. This means that, compared to other households, the price paid for the same good decreases for the low-income group.*

Figure 2.3: Estimated impulse responses of the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

The estimated responses for relative prices show a reversed picture compared to the paid price response. Even though only the impact response is significant for both groups, point estimates, as well as confidence bands for the low-income group, suggest that when one compares paid prices for identical goods, a monetary policy shock leads to comparatively higher prices for high-income households compared to low-income households. Additionally, taken together with the result that paid prices drop more for the high-income group in response to a monetary policy shock, this suggests that goods exclusively or primarily consumed by low-income

households drop less in price, and prices for those goods recover faster compared to goods exclusively or primarily consumed by high-income households. Furthermore, this indicates that the shock also leads to an adjustment of the price search behavior that is dissimilar between the groups. Since relative prices decline for the low-income group, this suggests that this group increases the expended price search effort compared to the high-income group.

To complete the discussion of household reactions, I will look at how the average product quality of the households' consumption basket[1] responds to the monetary policy shock. This will allow us to understand if the composition of the baskets changes and, additionally, if households substitute products for higher- or lower-quality products. Results are summarized in Fact 3.

Fact 3. (Product Quality Response) *A contractionary monetary policy shock leads to composition changes in the households' consumption baskets. In particular:*

1. *The average quality of the overall basket remains constant.*

2. *The average quality within departments significantly changes for some of the departments.*

The estimated response for the overall quality of households' consumption baskets suggests that, on average, there is no significant adjustment in product quality in response to a monetary policy shock. Notably, this does not directly imply that no systematic adjustments occur. Either different income groups react differently, and their adjustments countervail each other, or there are adjustments in different directions for various types of products, which cancel out in the overall consumption basket. Therefore, both avenues will be investigated in the following by first examining the responses for each product department separately and then estimating separate responses for high- and low-income households.

[1] As a brief reminder, the quality of a product is measured as the log-difference between the average price of a product and the average price of all products in the same module. Product quality is then aggregated using expenditure shares to construct a quality measure for a household's consumption basket.

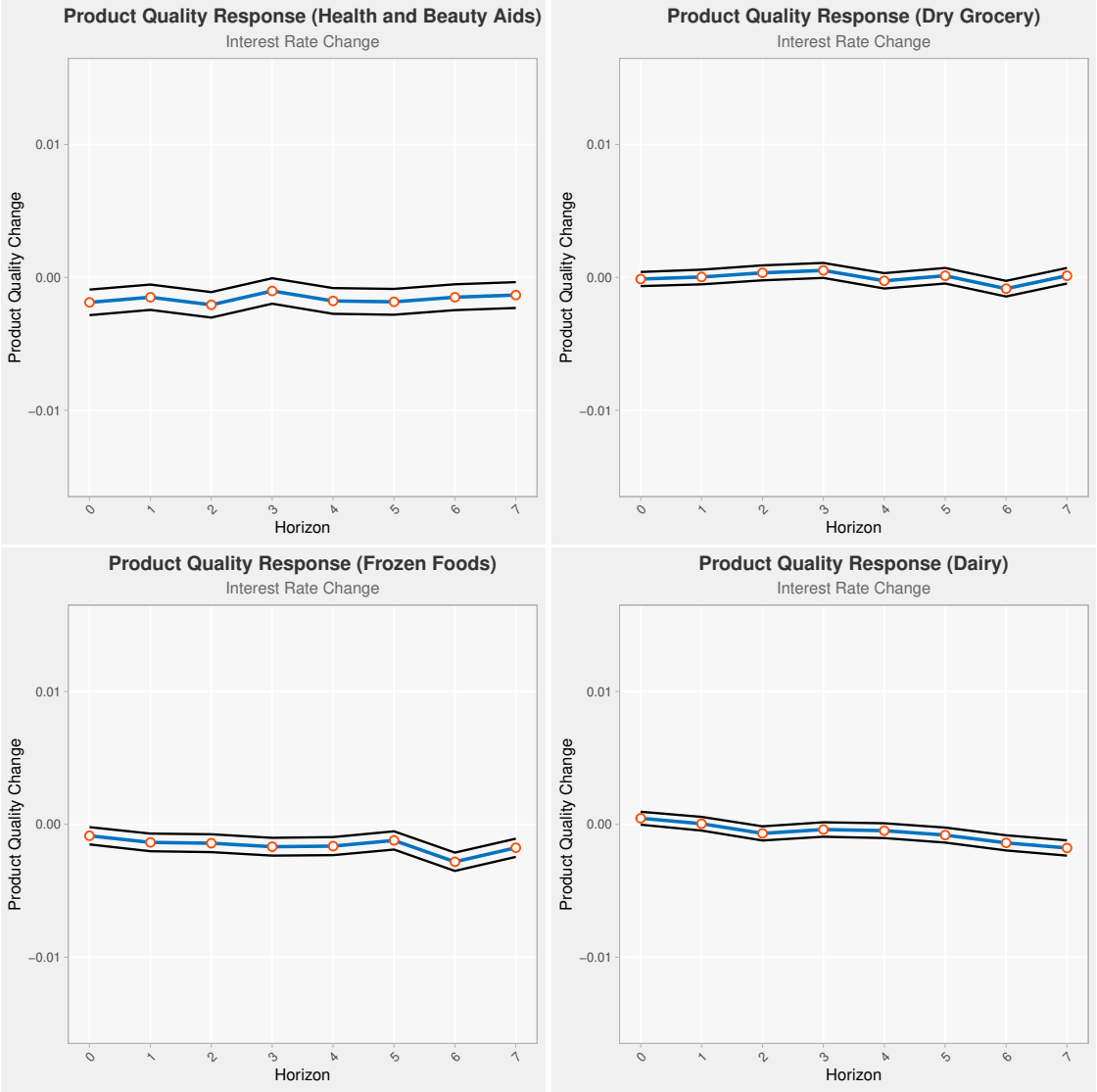


Figure 2.5: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

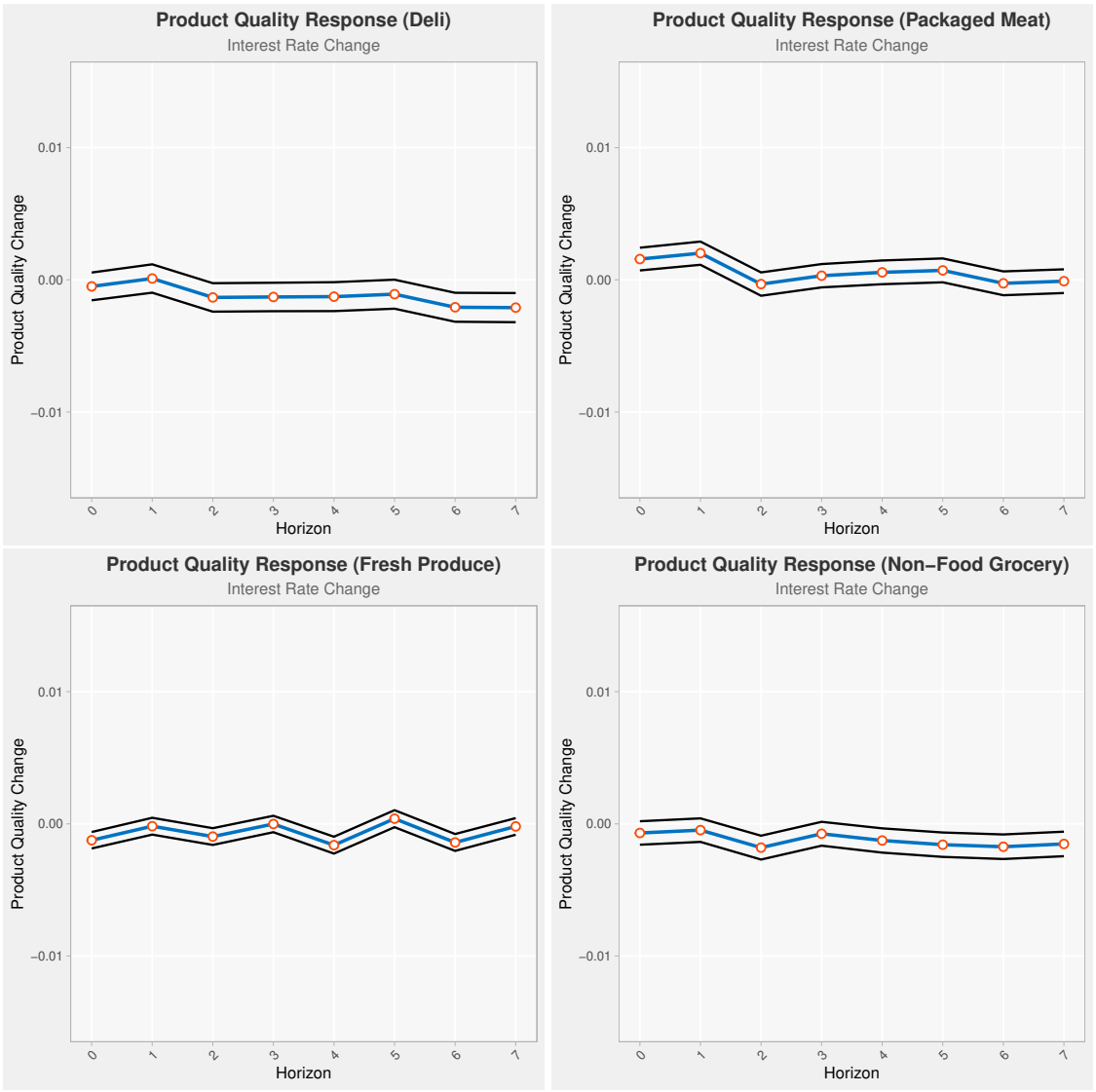


Figure 2.6: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

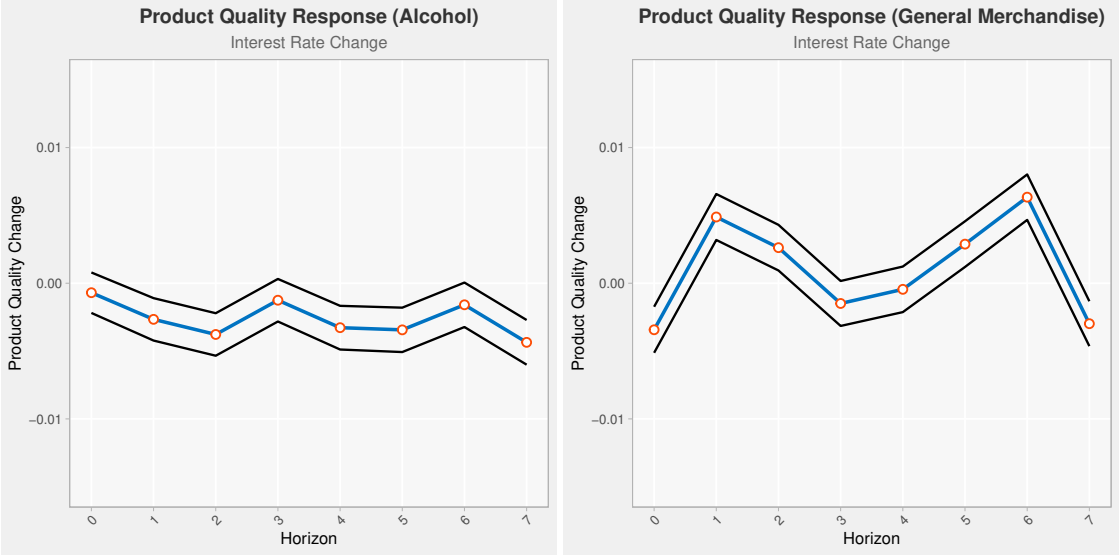


Figure 2.7: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Overall, the within-department quality responses show that there are clear adjustments in product quality within some departments. Some of the departments show positive quality adjustments, while others suggest a decrease in product quality or no change at all. This clearly indicates that the composition of the households' consumption basket is adjusted in response to the shock, in addition to the changes in the households' price search behavior documented above. The reason why these changes are not visible in the aggregate estimates is that the responses cancel out in aggregation. Even though for a substantial number of departments there is a negative quality response, this effect is canceled out by the few positive or close-to-zero responses. This happens because the same amount of money is not spent in each department. On the contrary, spending is distributed quite unevenly between departments.

Finally, I will briefly look at the responses for high- and low-income households. A full description is relegated to Appendix B.

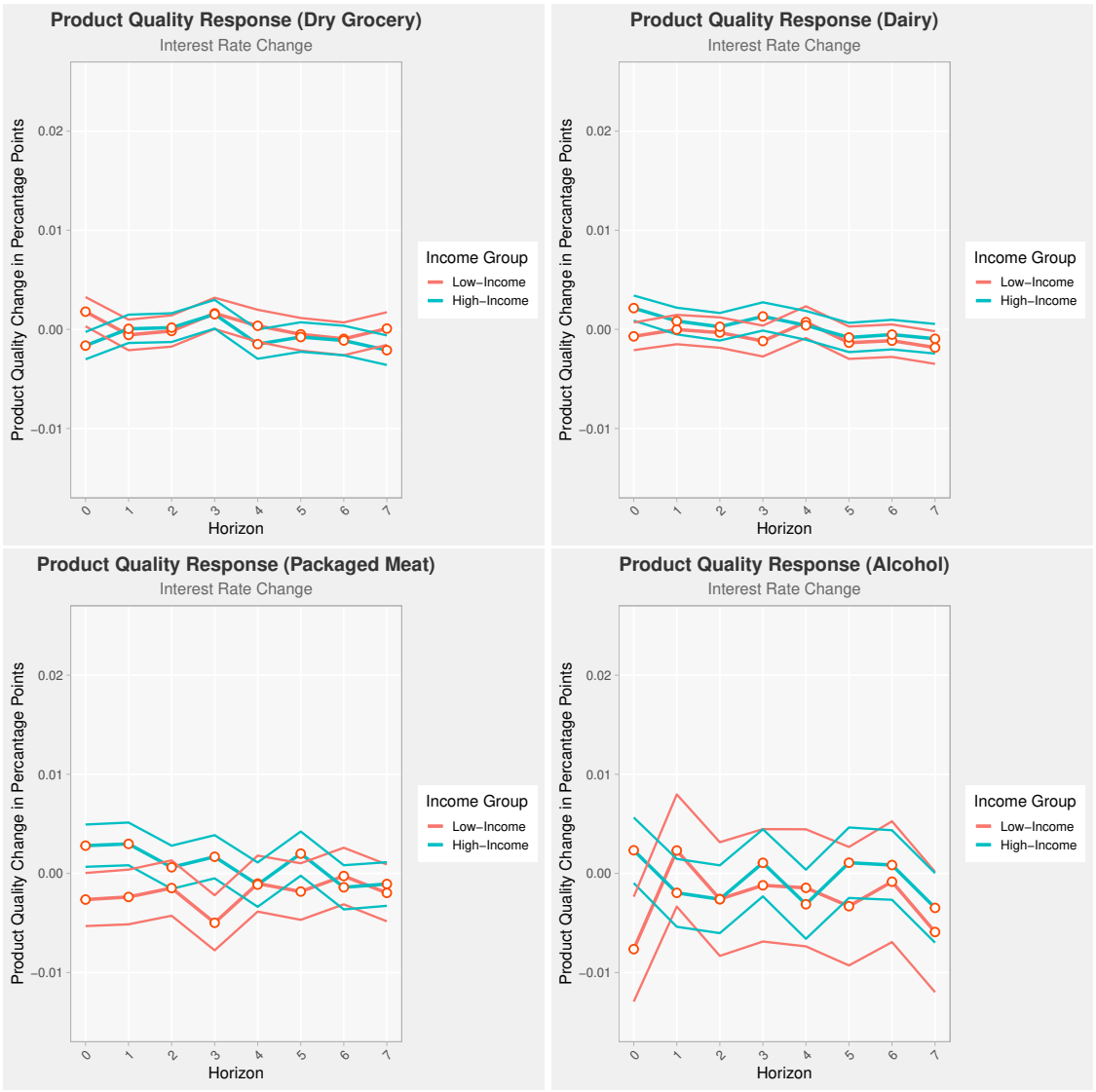


Figure 2.8: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

To briefly summarize the results, for aggregated product quality, splitting households by income and estimating separate responses for the top and bottom 20% does not substantially change the results. For within-department product quality, the results differ slightly. While for most departments the estimated responses are relatively similar, there are some where substantial differences exist between the responses, not only in magnitude but also in the sign of the response.

This clearly highlights the fact that not all households react in the same way to these shocks and that their response, at least partially, depends on their income level.

One possible mechanism for the different responses of households with different income levels is the following. Typically, differences in income are associated with differences in asset positions, especially when comparing the top and bottom 20% of the income distribution. While a household in the bottom 20% can be expected to have a negative asset position, a household in the top 20% is more likely to have a positive asset position. A change in the interest rate is likely to have different income effects on these two groups. Since one group is a net borrower and the other a net saver, the income effect will likely have opposite signs for these groups. This might at least partially explain the different adjustments in product quality.

III. Conclusion

While I have shown that the raw monetary policy shock has the expected effect on prices, this is not the case for the pure monetary policy shock. An undecomposed contractionary monetary policy shock leads to a decrease in prices of about 0.12% on impact, and the effect persists for about 7 quarters. For a contractionary pure monetary policy shock, I find a price increase of about 0.1%, with a peak of about 0.3% after 4 to 5 quarters. Quantitatively, the size of the response is in line with the inflation effect documented in Jarociński and Karadi (2020), although the sign for the pure monetary policy shock is different. Additionally, I find evidence that prices paid by high- and low-income households are affected differently. In response to a raw shock, products become relatively more expensive for high-income households compared to low-income households, while the difference is negligible for the decomposed shocks.

I showed that an undecomposed monetary policy shock leads to an increase in inflation after one quarter. For the pure monetary policy shock, the results are inconclusive. Additionally, in the aggregate, the average quality of households' consumption baskets remains unchanged, while quality changes are visible only within product departments. For some departments, there is a positive quality response, and for others, product quality decreases. Together with the fact that

both paid prices and relative prices change, this clearly indicates that households respond by changing both their price search behavior and the composition of the consumption basket.

Taken together, my findings show that there is indeed an effect of monetary policy shocks on paid prices and on the composition of households' consumption baskets, even though the effect is not always in line with conventional monetary theory. The fact that there is also an effect on relative prices suggests that there is a differential response along the income dimension. This could either mean that households adjust their search efforts differently depending on their income, or that goods consumed by both groups before the shock, which had a relatively high price differential, are either dropped by one of the groups or their share in the consumption basket decreases.

One interesting feature that emerges from this study is the counterintuitive effect of monetary policy shocks—specifically, that a contractionary shock leads to an increase in consumer prices for the pure monetary policy shock and in inflation for the undecomposed shock. The study therefore highlights the importance of adjustments in price-search behavior as well as in the composition of consumption baskets, factors that have mostly been absent from monetary policy models. While there has been some effort to construct models that include both price-setting frictions and price search (e.g. Burdett and Menzio, 2017), empirical evidence highlighting the importance of the price-search mechanism in this context has so far been lacking. This study fills that gap, while at the same time offering targets that might be useful for calibrating such models and pointing to avenues for further research. A promising avenue for future research could involve developing a monetary policy model that specifically addresses these findings, especially given the observation that relative prices and basket composition also change in response to such shocks.

Bibliography

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

Argente, David and Munseob Lee (2021) "Cost of Living Inequality during the Great Recession," *Journal of the European Economic Association*, 19 (2), 913–952.

Burdett, Kenneth and Kenneth L Judd (1983) "Equilibrium Price Dispersion," *Econometrica*, 51 (4), 955–69.

Burdett, Kenneth and Guido Menzio (2017) "The (Q, S, s) Pricing Rule," *The Review of Economic Studies*, 85 (2), 892–928.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (1999) "Chapter 2 Monetary policy shocks: What have we learned and to what end?", 1 of Handbook of Macroeconomics, 65–148: Elsevier.

D'Acunto, Francesco, Ulrike Malmendier, Juan Ospina, and Michael Weber (2021) "Exposure to Grocery Prices and Inflation Expectations," *Journal of Political Economy*, 129 (5), 1615–1639.

Gertler, Mark and Peter Karadi (2015) "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7 (1), 44–76.

Jarociński, Marek and Peter Karadi (2020) "Deconstructing Monetary Policy Surprises— The Role of Information Shocks," *American Economic Journal: Macroeconomics*, 12 (2), 1–43.

Kaplan, Greg and Guido Menzio (2015) "The Morphology of Price Dispersion," *International Economic Review*, 56 (4), 1165–1206.

Kaplan, Greg and Sam Schulhofer-Wohl (2017) "Inflation at the Household Level," *J. Monet. Econ.*, 91, 19–38.

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

Pytka, Krzysztof ⓡ Daniel Runge (2025) "Looking into the Consumption Black Box: Evidence from Scanner Data," Working paper.

Ramey, V.A. (2016) "Chapter 2 - Macroeconomic Shocks and Their Propagation," 2 of Handbook of Macroeconomics, 71–162: Elsevier.

Romer, Christina D. and David H. Romer (2004) "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94 (4), 1055–1084.

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

2.A. Appendix

A. Responses for Pure Monetary Policy and Central Bank Information Shock

This section contains the additional results for the decomposed shocks. It follows the same structure as the main part, starting with the effect on the average paid price, then relative prices, inflation and finally the next part contains the product quality responses.

I want to check how both paid prices as well as relative prices respond to the pure monetary policy shock and the central bank information shock.

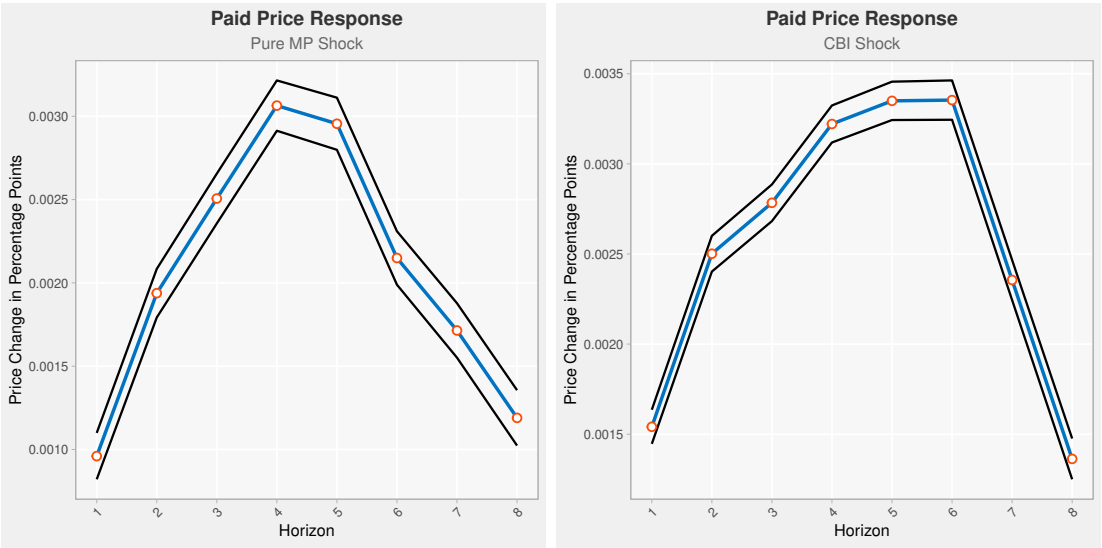


Figure 2.9: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Starting with the specification where all products are weighted equally, the estimated impulse responses are quite surprising. In contrast to the analysis in Jarociński and Karadi (2020), both shocks lead to an increase in prices. Both responses are significant over the entire estimated horizon and feature no sign changes. The magnitude of the effect is identical to the estimated response for the undecomposed shock. For the pure monetary policy shock, the peak response is reached after four quarters, and for the information shock, at quarter six. In both cases, the response then reverts back toward 0. As can be seen in the following plots for the two decomposed shocks, there is no substantial difference when weighting products by expenditure shares.

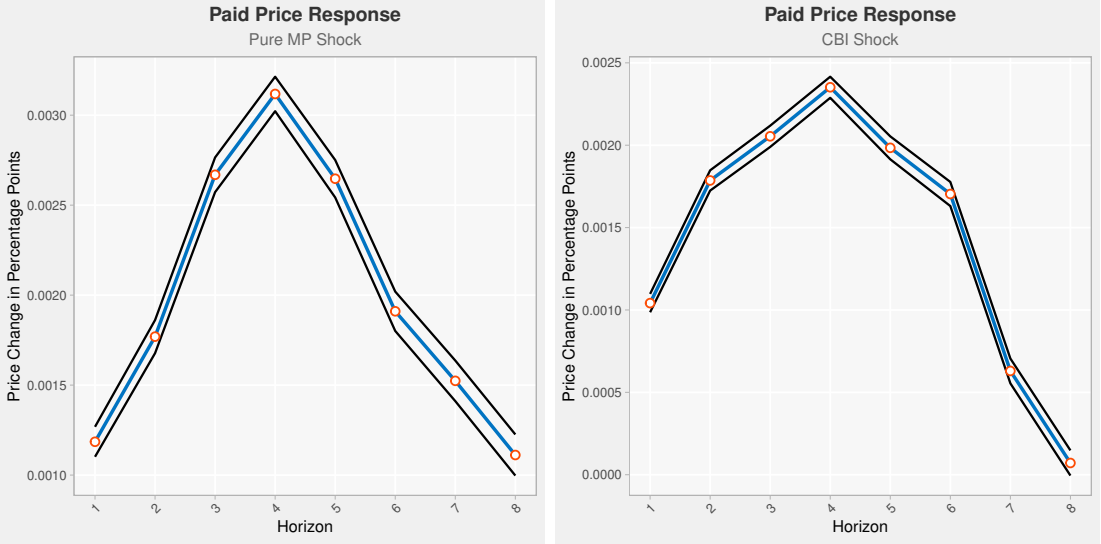


Figure 2.10: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

For the equally weighted specification, estimating the responses split by income group reveals almost no additional information. The one major difference is that the responses drop back to 0 faster after their peak. For the expenditure-weighted specification, there is a significant difference in responses between the two groups. The prices paid by the low-income group respond more strongly to the pure monetary policy shock, while the central bank information shock leads to a stronger price response for the high-income group.

Figure 2.11: Estimated impulse responses for the top and bottom 20 % of the income distribution including 95% confidence bands. All products are weighted equally. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Figure 2.12: Estimated impulse responses for the top and bottom 20 % of the income distribution including 95% confidence bands. All products are weighted using expenditure shares. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

For relative prices, the overall picture is similar for the response to the raw monetary policy shock for both the pure monetary policy shock and the central bank information shock. In both cases, there is a tendency that goods become relatively less expensive for the low-income group, even though the difference between the two responses, as well as the responses themselves, are not always significant. Notably, for the pure monetary policy shock, the response of the high-income group is significantly positive over the first four quarters.

Figure 2.13: Estimated impulse responses top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Next, I estimated the response of the inflation index to a monetary policy shock. The results are summarized in Fact 4.

Fact 4. (Inflation Response) *A contractionary monetary policy shock leads to an increase in the inflation rate. In particular:*

1. *Inflation drops by about 0.1% on impact.*
2. *Besides the impact response, inflation increases permanently.*
3. *There is no economically significant difference between the inflation response of the top and bottom 20% of the income distribution.*

From the estimated response, we can see that while there is a small but negative impact response, a contractionary shock leads to an increase in inflation after just one period, which persists over the entire estimated horizon. This is strong evidence for the presence of a price puzzle within the inflation rate response to a monetary policy shock when using paid price data. Interestingly, in Jarociński and Karadi (2020), the responses of the GDP deflator generated from the same set of shocks do not show a price puzzle. There are three potential causes for the difference between the estimated response of inflation here and the response of the GDP deflator in Jarociński and Karadi (2020). First of all, it could be due to using paid prices within the calculation and therefore capturing more of the household response. Second, it could be due to the fact that consumption baskets are allowed to change from period to period, while the deflator is computed from nominal and real GDP. Finally, it might be due to the fact that inflation here is only calculated for a subset of overall consumption due to the limitations that come with the dataset.

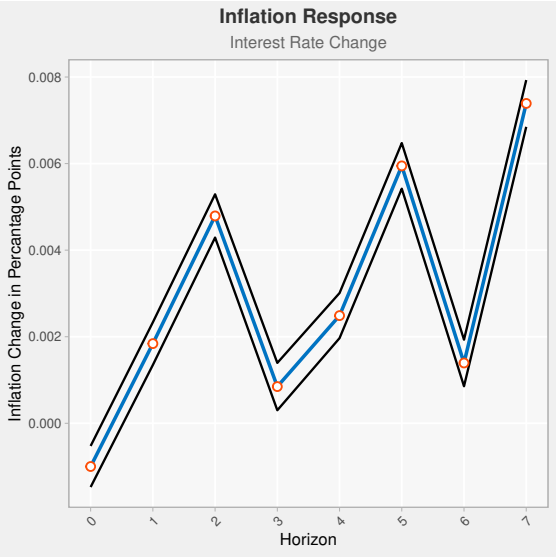


Figure 2.14: Estimated impulse response for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Besides that, the very different response in inflation and average paid prices seems puzzling. One potential reason for this is the way the index is constructed

and the fact that this excludes a sizable number of products for a significant number of households. As shown in Pytka ⓡ Runge (2025), the average likelihood that a given product stays within a household's consumption basket is very low. Since the inflation index only includes those products that a given household purchases in a given quarter and then repurchases in the same quarter the next year, a high number of products are excluded for most households. Given the fact that higher expenditure goods seem to respond more inflationary than lower expenditure goods, and that persistence—i.e., the likelihood of leaving the consumption basket—is higher in terms of expenditures, this might bias the estimated inflation response in a more inflationary direction.

Finally, for inflation, the response to the pure monetary policy shock is inconclusive. There is a short increase in inflation after an insignificant impact response, which is directly followed by a decrease. For the CBI shock, the results are more clear. A contractionary CBI shock leads to a persistent decrease in inflation, after an inflationary impact response. The peak response is an inflation decrease of 0.5%.

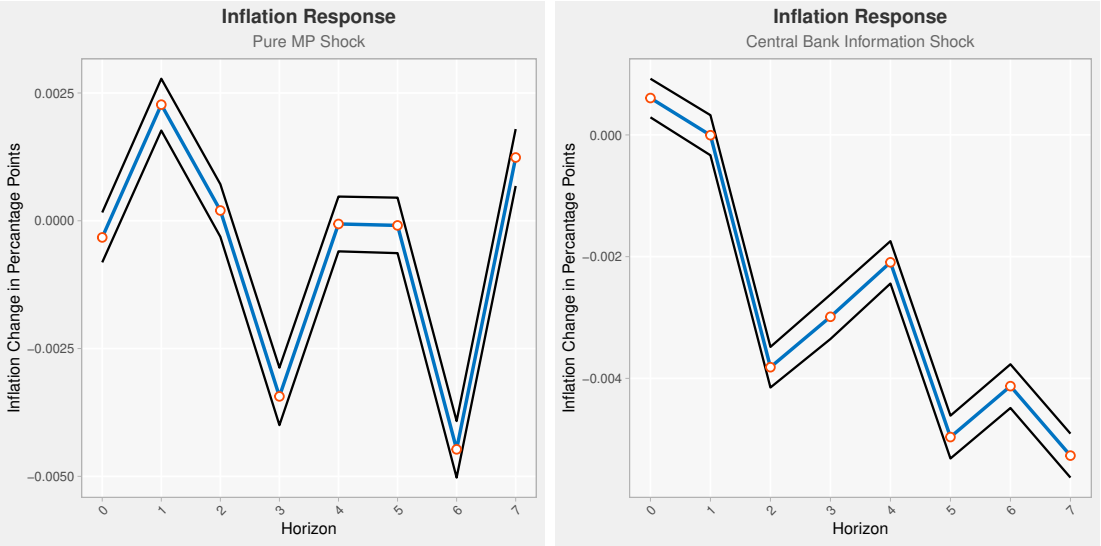


Figure 2.15: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Next, I estimate impulse responses for the top and bottom 20% of the income

distribution separately to check if there are differences in inflation responses along the income distribution. The overall shape of the responses is very similar to the estimates for all households jointly, suggesting that there are no major differences in the way the two income groups respond to the shocks. Additionally, the differences between the two groups themselves are mostly insignificant. When comparing the point estimates, it seems that overall, the low-income group responds slightly more strongly to the shocks than the high-income group, although these differences are mostly insignificant.

Figure 2.16: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Also for the two decomposed shocks reestimating the responses separately for the top and bottom 20% of the income distribution does not reveal additional insights.

Figure 2.17: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

B. Quality Responses

This section contains the responses for within-department product quality to the two decomposed shocks as well as the within-department responses estimated separately for high- and low-income households for all three shocks. Instead of describing each response in detail, I will highlight the most pronounced features.

Within-Department

Starting with the response for the pure monetary policy shock, we can classify the departments into roughly one of three categories: responses are predominantly positive or insignificant, predominantly negative or insignificant, or overall insignificant or inconclusive. For Health and Beauty Aids, Dairy, Deli, and Packaged Meat, responses are predominantly positive, while for Frozen Foods, Non-Food Grocery, Alcohol, and General Merchandise, responses are predominantly negative. Finally, for Dry Grocery, Frozen Foods, and Fresh Produce, results are inconclusive.

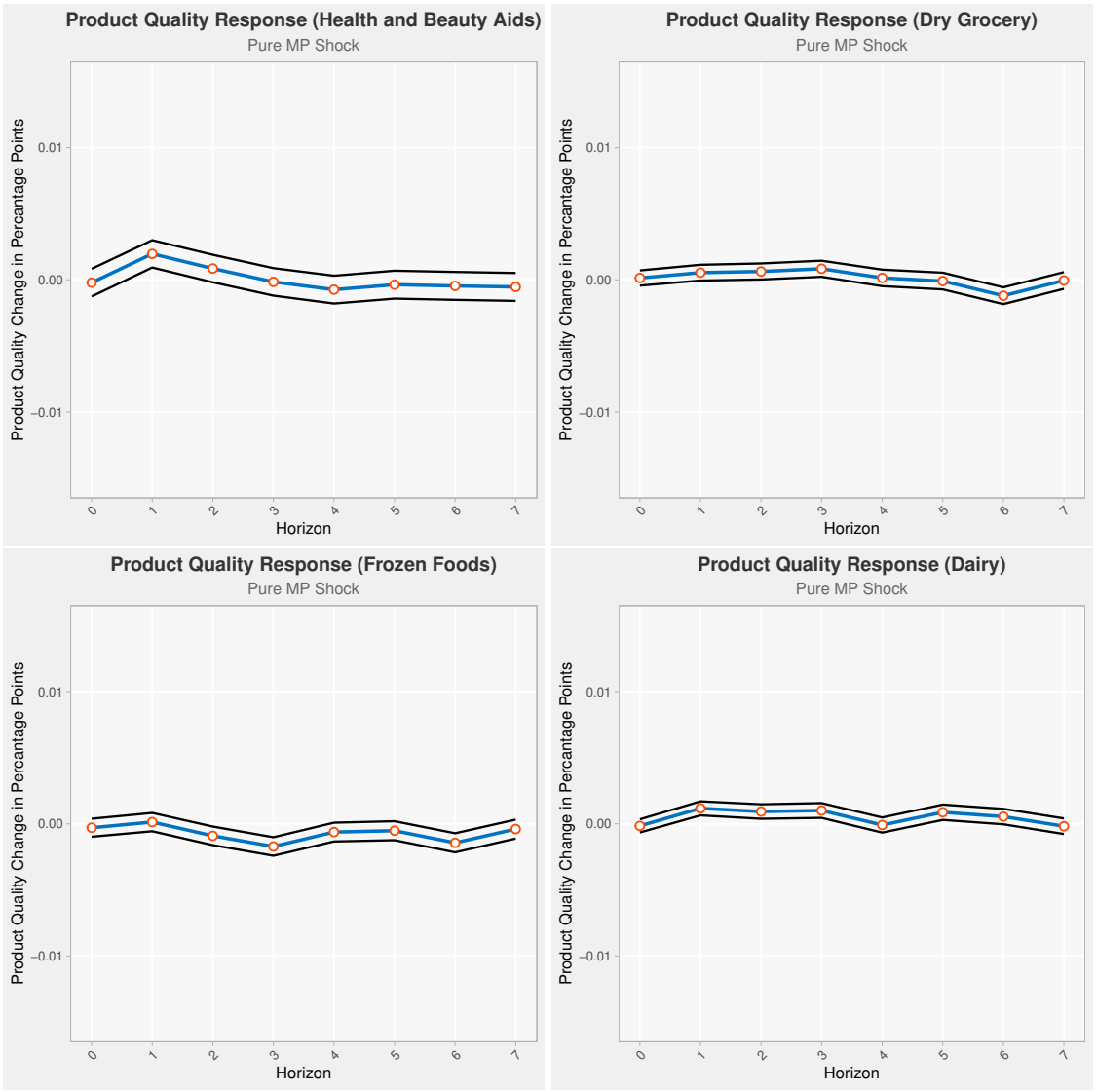


Figure 2.18: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

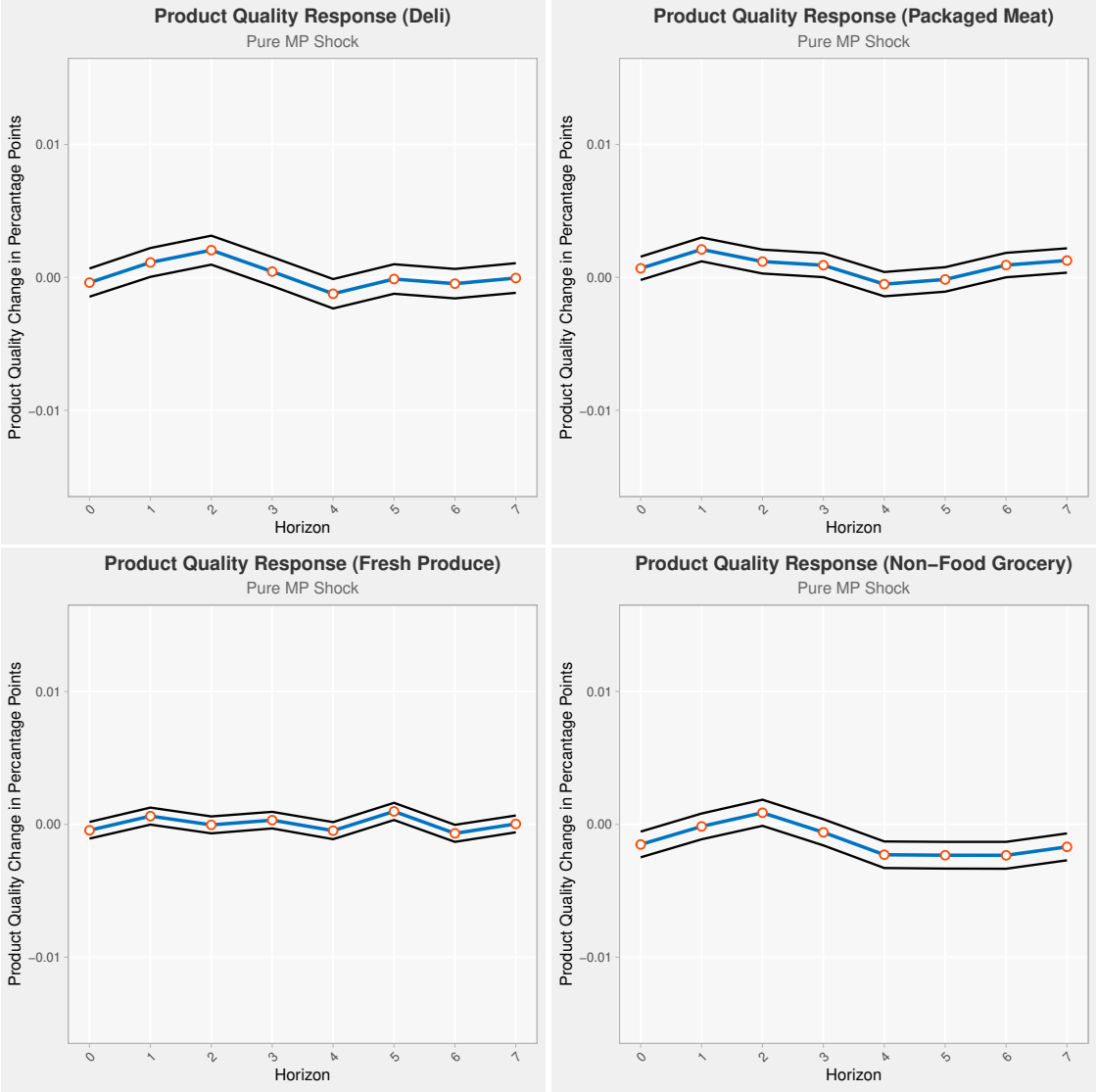


Figure 2.19: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

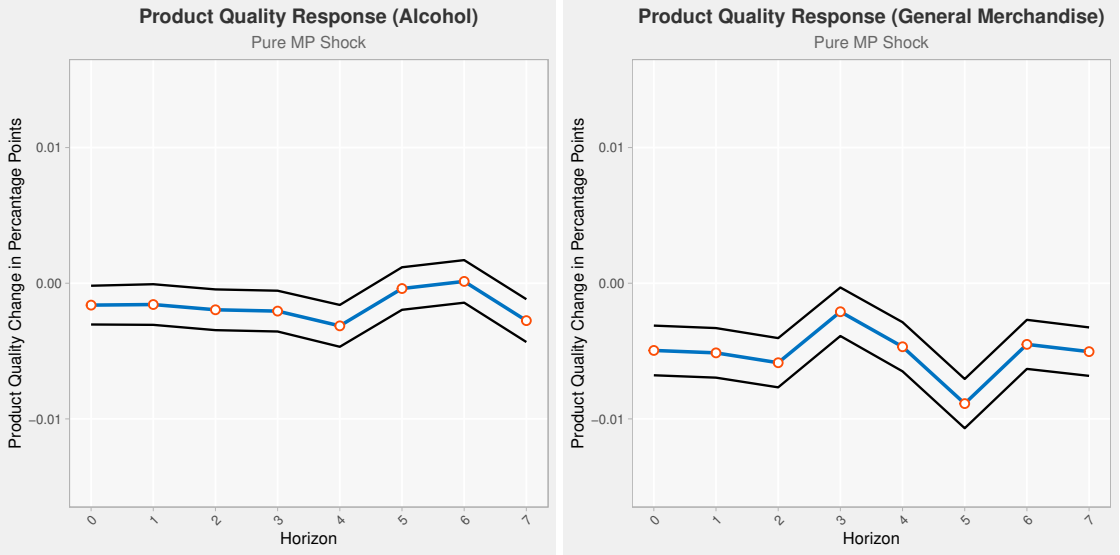


Figure 2.20: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

For the central bank information shock, we observe a predominantly positive response only for General Merchandise. For Health and Beauty Aids, Frozen Foods, Dairy, Deli, Packaged Meat, Fresh Produce, and Alcohol, the response is predominantly negative. Finally, an inconclusive response is observed for Dry Grocery and Non-Food Grocery. This clearly shows that the two decomposed shocks affect average product quality differently within the individual departments.

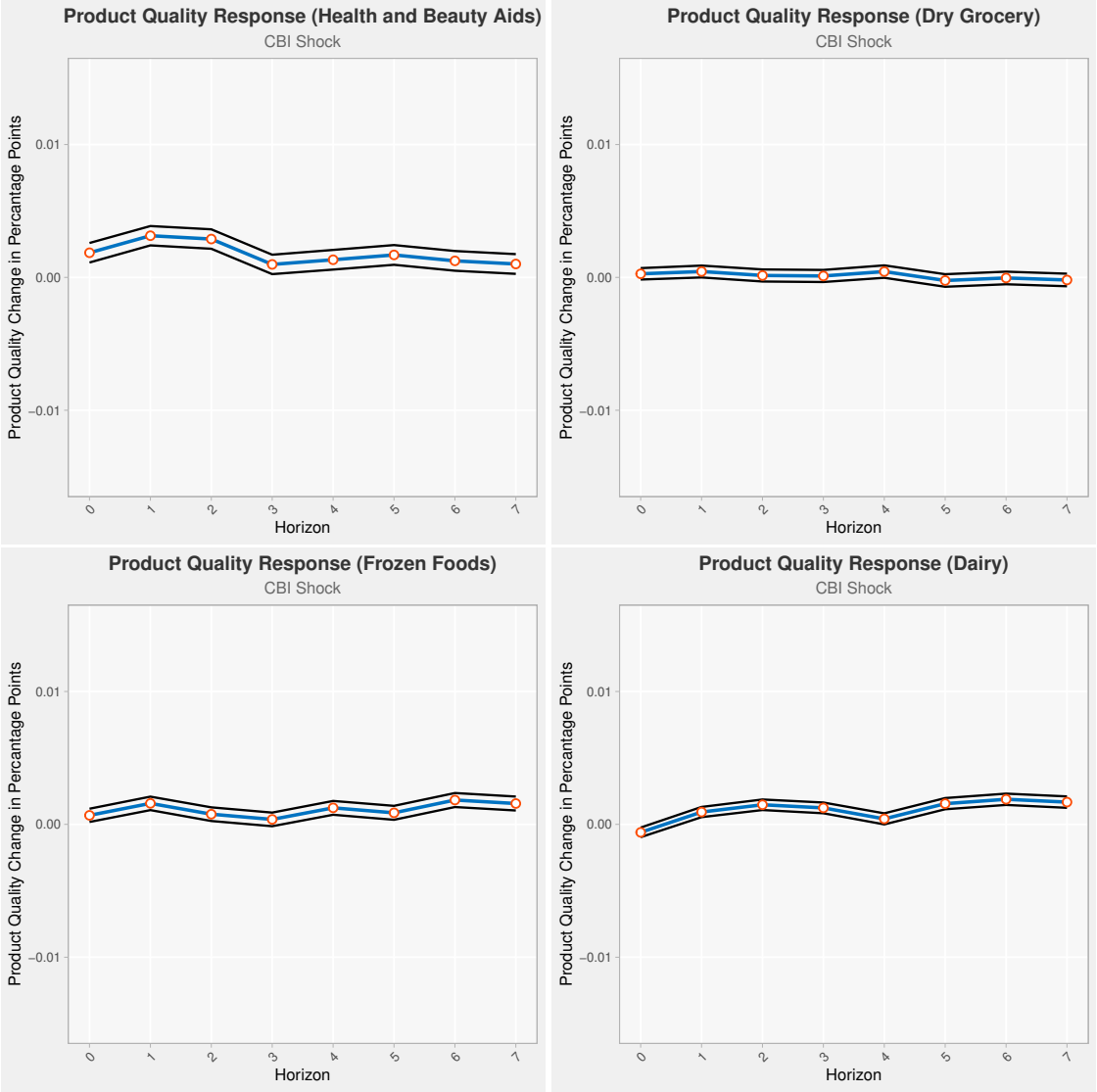


Figure 2.21: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

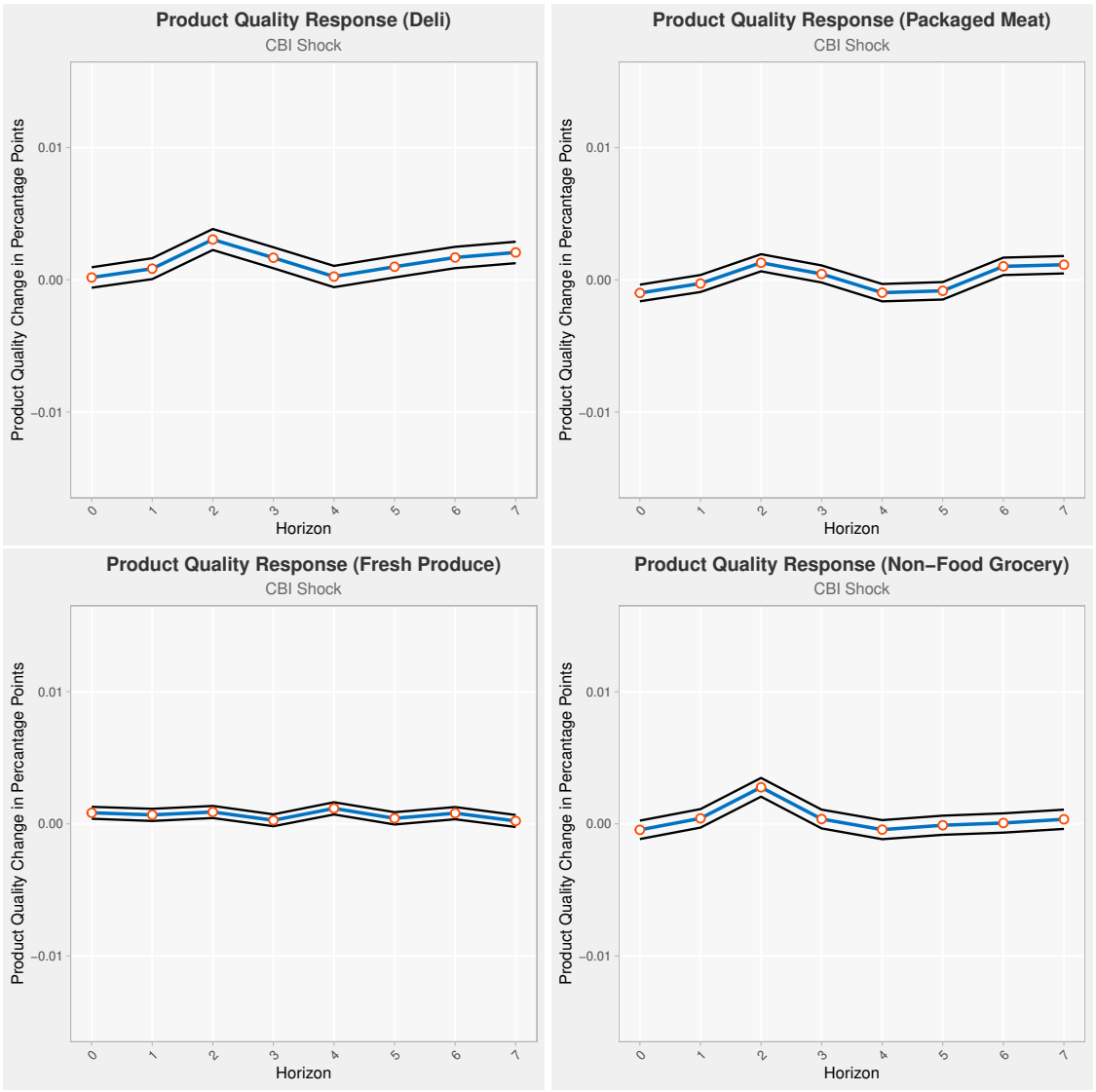


Figure 2.22: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

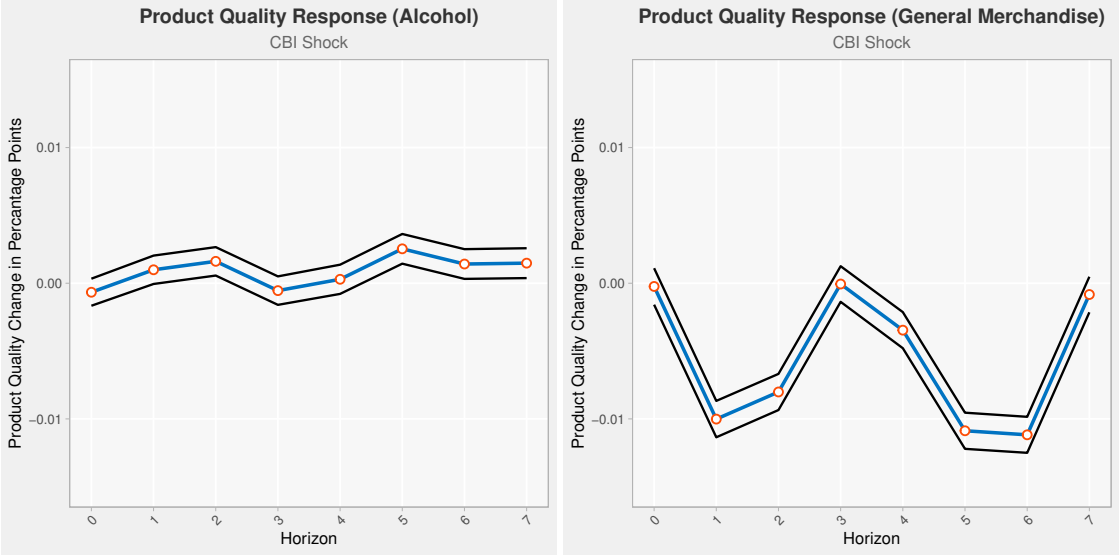


Figure 2.23: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

All in all, the estimated responses show that significant quality adjustments occur within the different departments for all three shocks. Although the overall quality of the consumption basket does not change significantly, as indicated by the insignificant responses for overall product quality, substantial adjustments take place. This suggests that households respond to the shocks by altering the composition of their consumption baskets and adjusting quality choices within the different departments. One potential cause for these differential responses might be that the price-quality ratio changes differently across departments in response to the shocks. This hypothesis, however, cannot be easily investigated within this study, as product quality is not observed directly.

Within-Department split by Income

The estimated responses of aggregate product quality for the top and bottom 20% of the income distribution show no significant responses for both of the groups. Noticeably, the point estimates for the low-income group are less volatile than those of the high-income group.

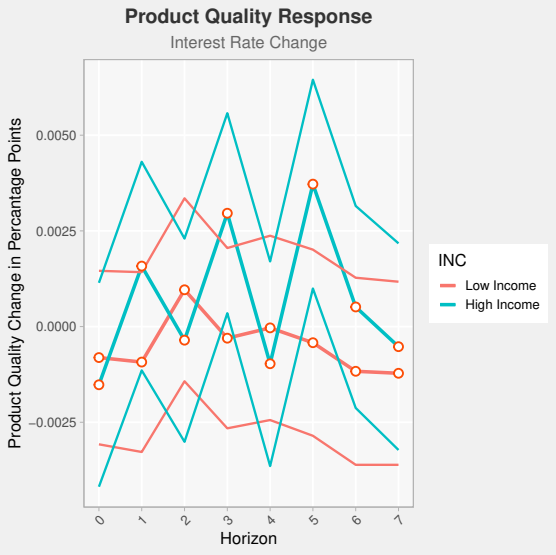


Figure 2.24: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

For the pure monetary policy shock, both impact responses are insignificant. For the low-income group, the responses are insignificant up to quarter 2, where the response becomes significantly positive. After that, it drops again and becomes insignificant for the rest of the estimated horizon. For the high-income group, the estimated responses are insignificant except for quarters 3 and 5, where the response is significantly positive. As with the interest rate shock, volatility is higher for the high-income group.

For the central bank information shock, the estimated responses of the high-income group are insignificant throughout the entire time period. For the low-income group, the responses are significantly negative up until quarter 3, after which they turn insignificant. Overall, the estimated responses show that while the overall reaction to the shocks is similar across the two income groups, there are some significant differences. The low-income group appears to respond with less volatility.

Figure 2.25: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Lastly, I will look at the responses of within-department product quality separately for high- and low-income households. Here, I will only highlight those departments where there are significant differences between the two income groups.

For the undecomposed shock, we mostly observe differences in the impact response. For Health and Beauty Aids, the response for the high-income group is significantly negative, while being insignificant for the low-income group. For Dry Grocery, the response is significantly positive for the low-income group and negative for the high-income group. In the Dairy and Packaged Meat department, only the high-income group's response is significantly positive. For Alcohol, only the low-income group's response is significantly negative, while for General Merchandise, only the high-income group's response is significant and negative.

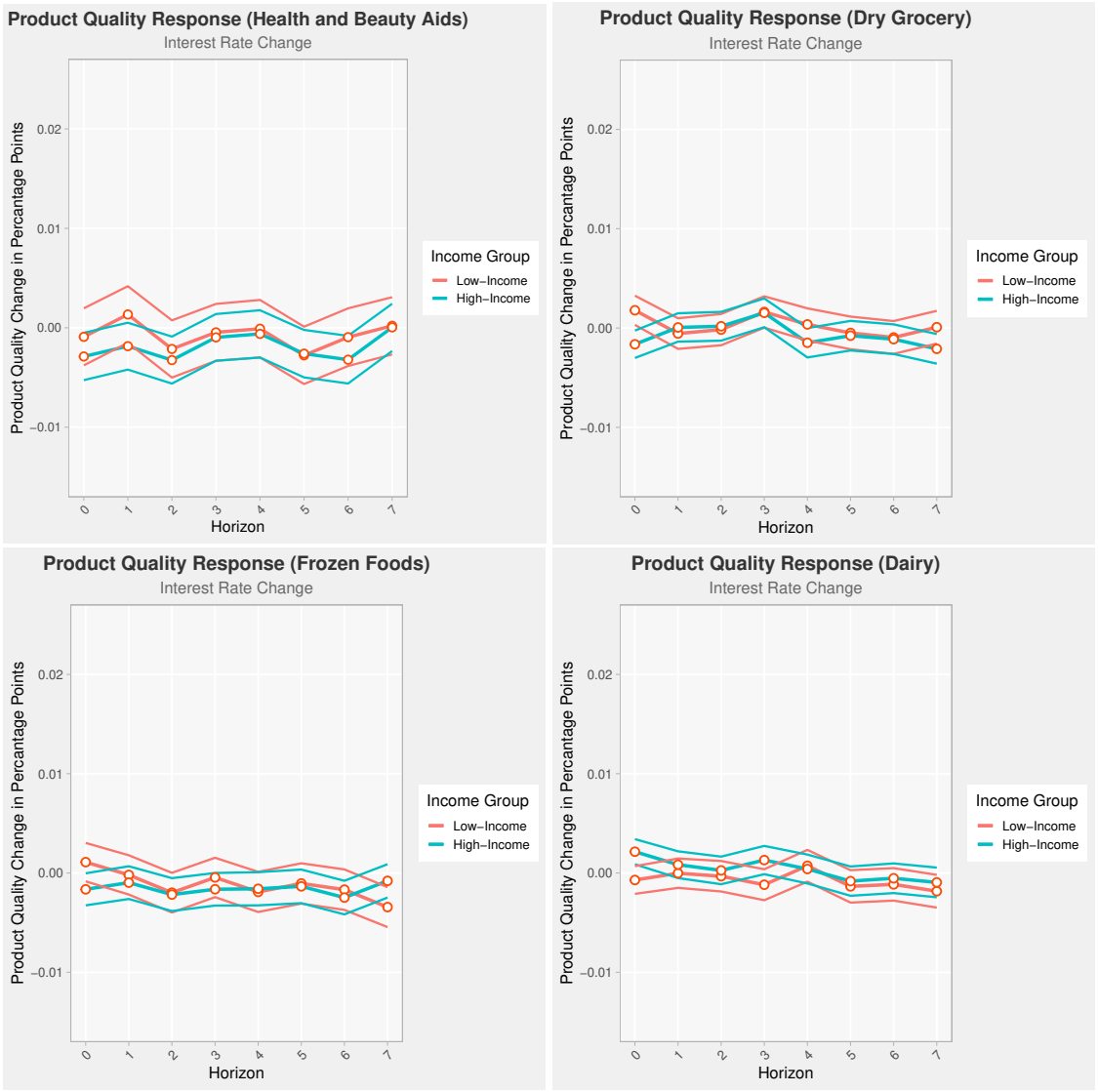


Figure 2.26: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

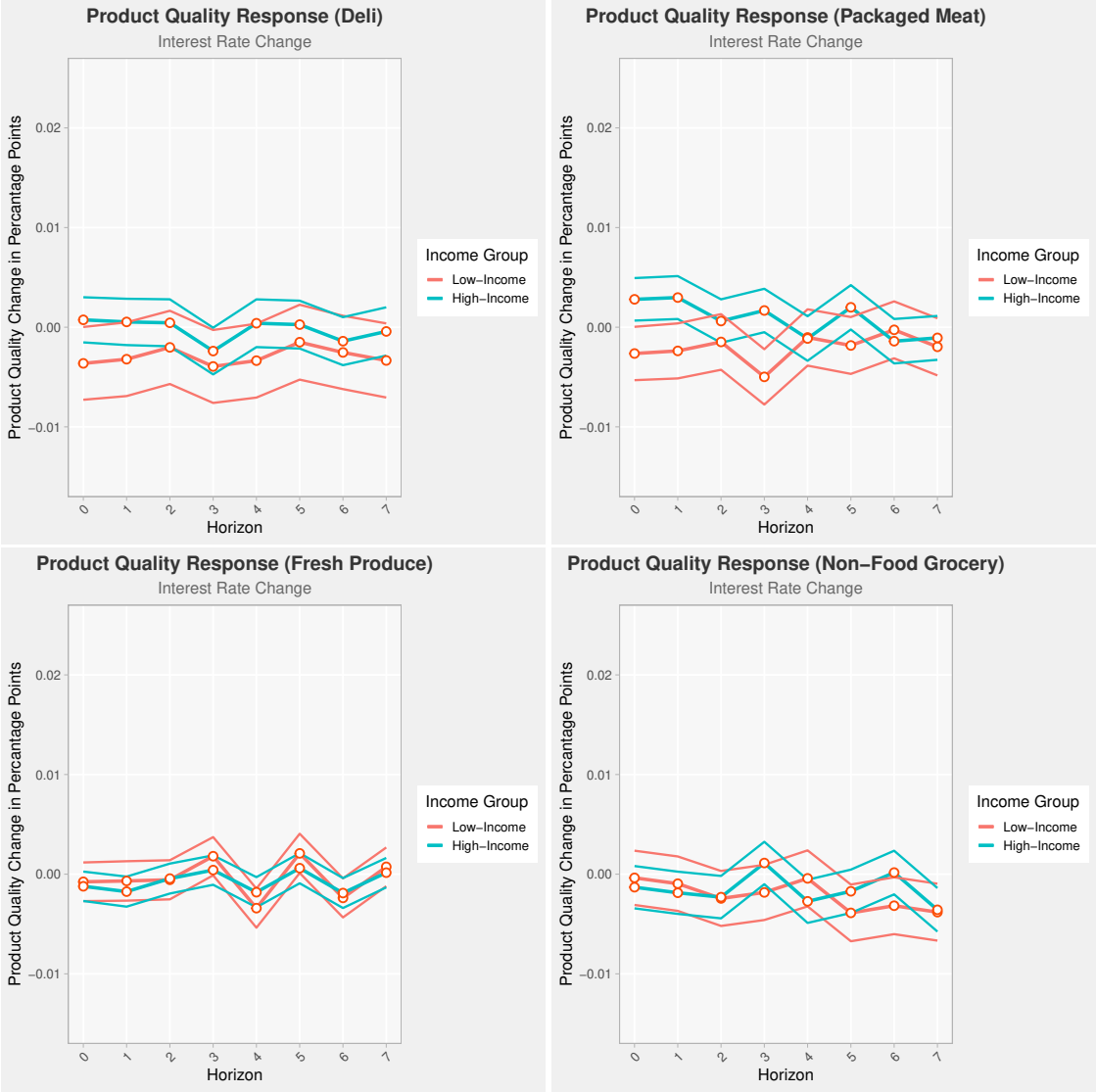


Figure 2.27: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Figure 2.28: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

For the pure monetary policy shock, for both Health and Beauty Aids and Dry Grocery, only the low-income group's response is significant and positive. For the Alcohol department, the response of the low-income group is significant and positive.

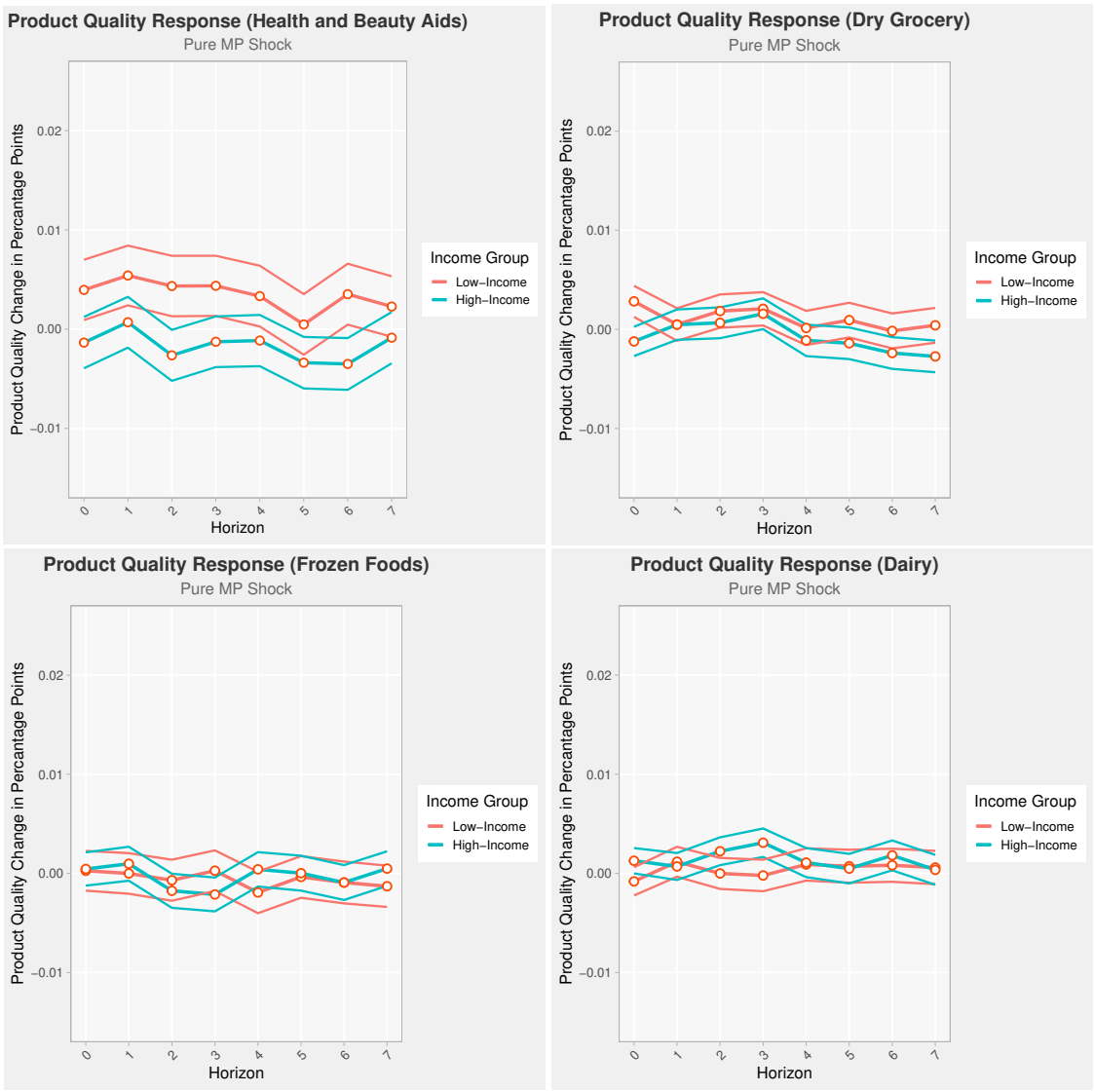


Figure 2.29: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

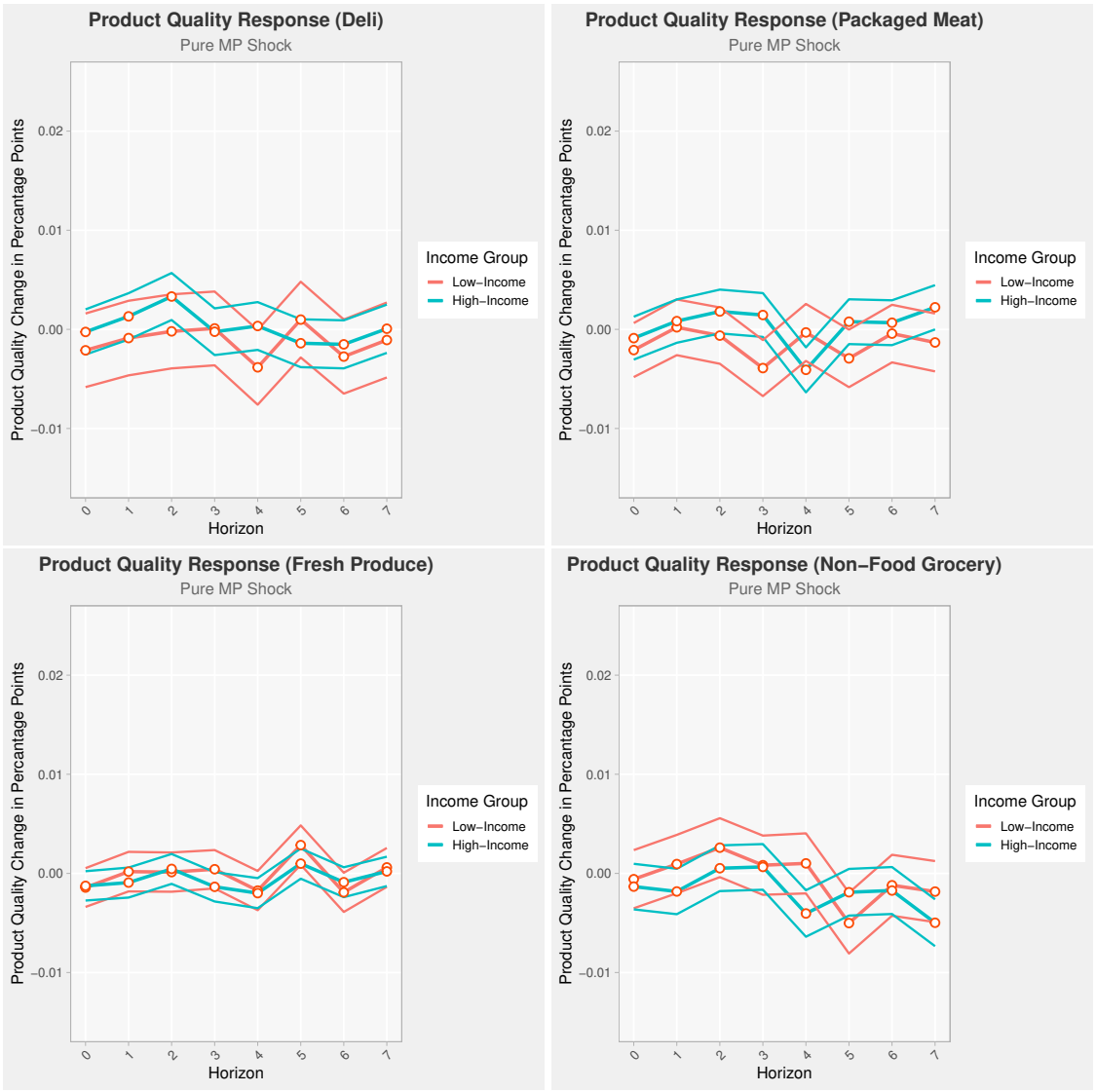


Figure 2.30: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Figure 2.31: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Finally, for the central bank information shock, the low-income group's response for the Health and Beauty Aids department is significantly negative. For Frozen Foods, only the high-income group's response is significant and negative. The response of the high-income group for Packaged Meat is significant and positive. Finally, for the Alcohol department, the high-income group's response is significantly positive, and for General Merchandise, it is significantly negative.

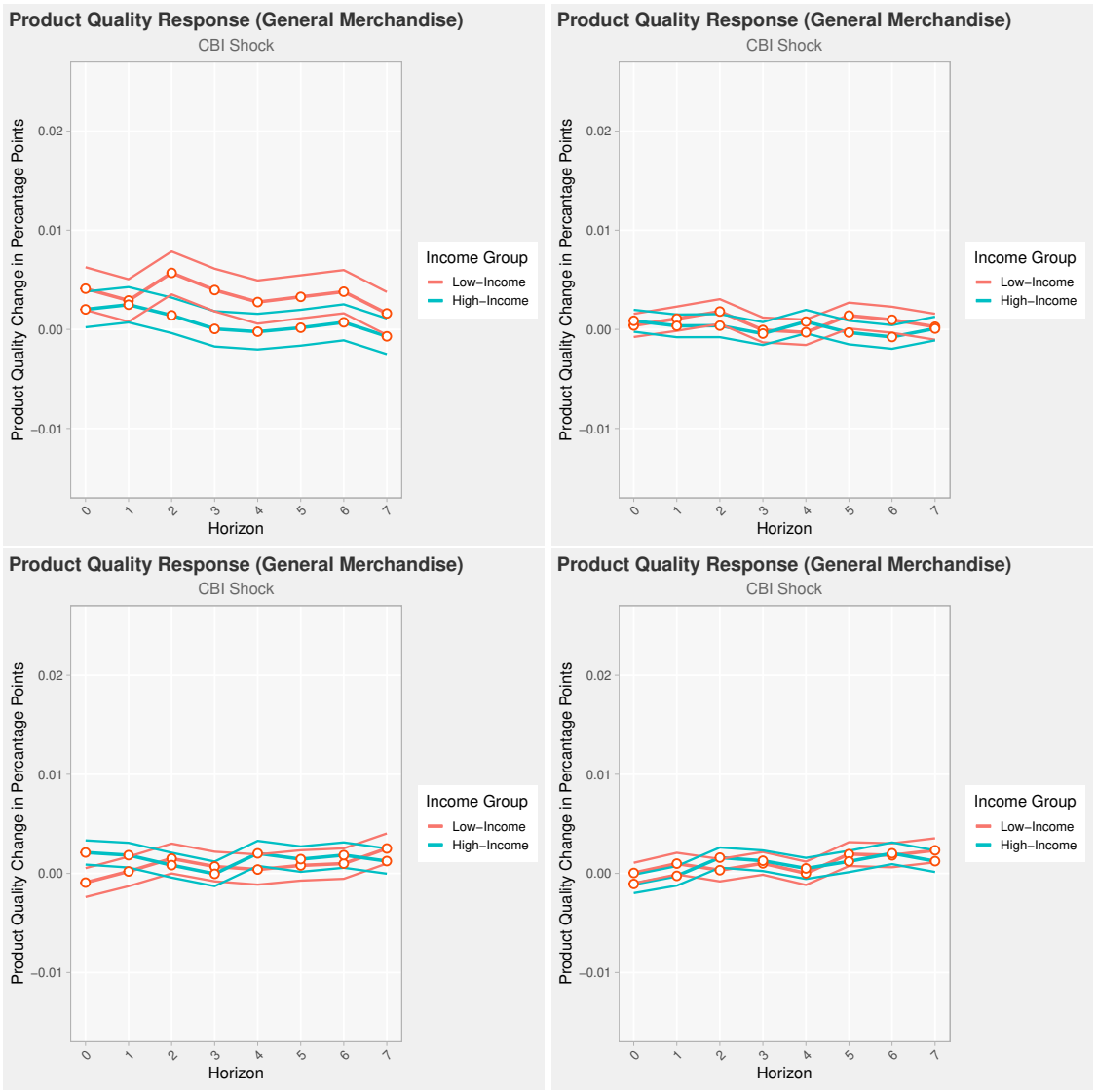


Figure 2.32: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

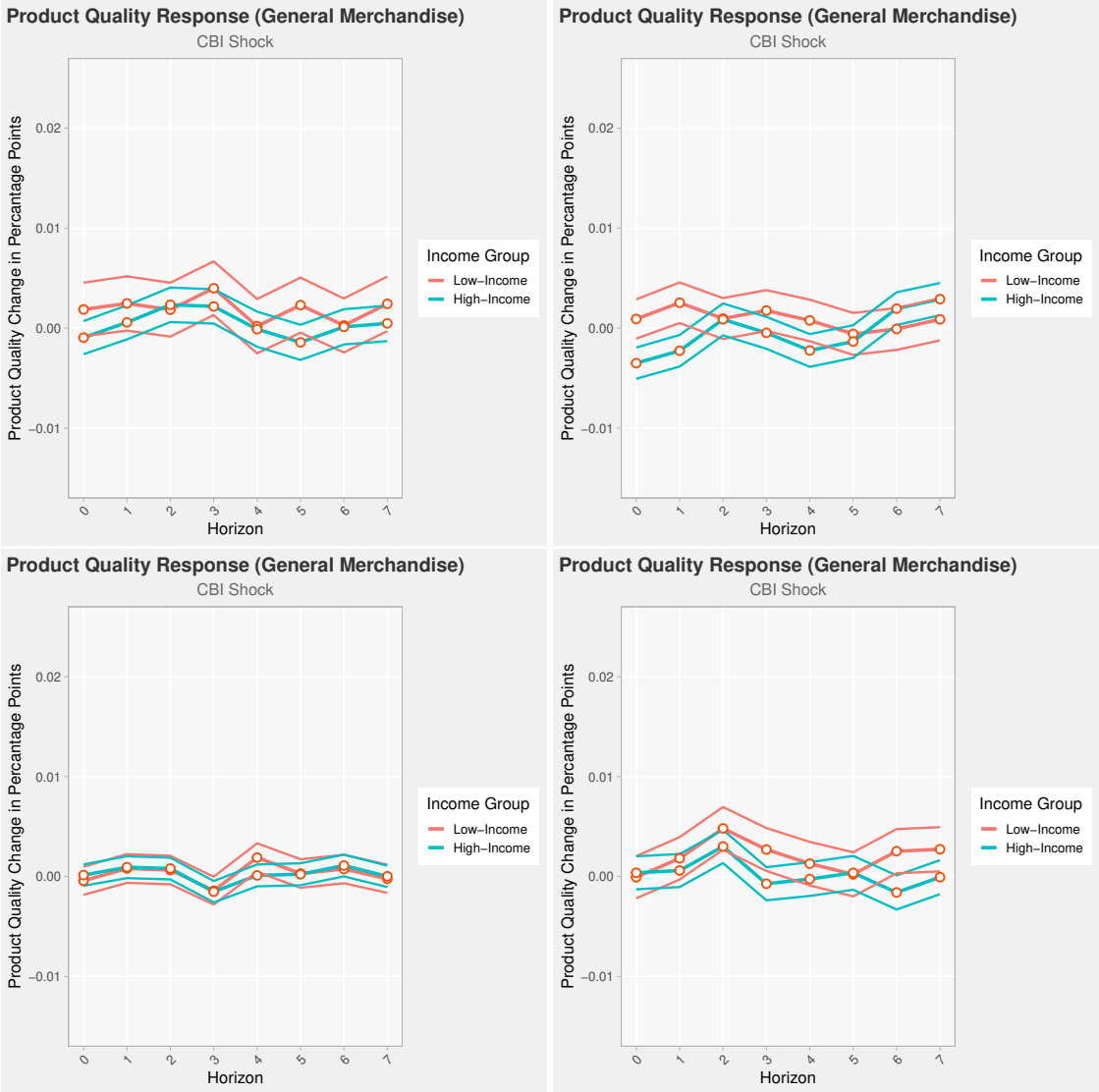


Figure 2.33: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Figure 2.34: Estimated impulse responses for the top and bottom 20% of the income distribution including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

All in all, the estimated responses show that how households' product quality choices respond to monetary policy shocks depends, to some extent, on their income level. While responses are similar across many departments, there are some in which low-income households respond significantly differently compared to those of high-income households. Not only is the size of the response different, but in some cases, the direction is even opposed. This clearly highlights that not all households are affected similarly by monetary policy shocks, which might be at least partially explained by the varying effects of the shock on household wealth, depending on their asset position.

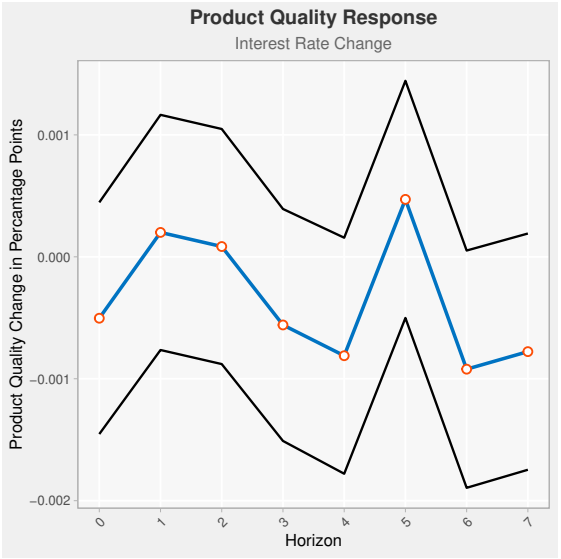


Figure 2.4: Estimated impulse responses for all households within the sample including 95% confidence bands. The plotted response is normalized to show the effect of a contractionary one standard deviation shock.

Chapter 3

The Mirage of Consumption Sorting: How Data Sparsity and Search Frictions Create Spurious Quality Ladders

This paper explores how, when quality is measured using a product's price, data sparsity and search frictions can lead to artificial quality differences between products. Using a search-theoretical model, I demonstrate that even when products are identical, sparsity combined with price search frictions can create spurious quality ladders. Furthermore, it can lead to falsely concluding that higher-income households consume higher-quality goods. In addition, using data from the Kilts-NielsenIQ Consumer Panel (KNCP), I find that quality differences appear smaller for products with more purchases, suggesting that indeed small sample bias plays a role in the quality estimation.

I. Introduction

Data sparsity has been an issue in many contexts involving high-dimensional datasets. This means that if one imagines a choice matrix containing all choices made by the entities, the most common element in this matrix will be zero, or, put otherwise, a most choices are much more likely not to be chosen than to be chosen. Especially whenever the set of possible choices is much larger than the number of entities observed making these choices it is quite likely that a high degree of sparsity will lead to spurious results, if not properly accounted for.

This problem has been accounted for in some recent studies, for example Gentzkow et al. (2019) in the context of political speech and Pytka ⓡ Runge (2025) in the context of consumption choices, however it is not yet common practice to explicitly address issues of sparsity. The potential issues that can arise from sparsity become clear quite quickly when considering some simple thought experiments. When looking at a sparse dataset, there are potentially a considerable number of choices that are only chosen by a very small number of entities and - in extreme cases, only a single entity. The fact that these choices are not observed more frequently within the dataset could have two distinct explanations with very different implications. The first possibility is that these choices are indeed relatively unpopular among the entities, and even if we observed a significantly higher number of entities, the choices would still be observed relatively rarely. In this case, ignoring sparsity would not be an issue.

The other possibility is that these choices are similarly popular to the ones we observe being made more often, and the only reason we observe them relatively rarely is the limited size of the dataset and the fact that the number of possible choices is significantly higher than the number of entities. In this case, increasing the number of observed entities would change the conclusion about the popularity of the choice. A naive estimation approach would treat both cases the same and potentially attribute a high degree of explanatory power to these cases.

In this project, a very similar problem to the one outlined above is addressed. In the empirical literature about consumption, a sizable number of studies have been concerned with the issue of measuring product quality. Since the quality of a product cannot be observed directly, it is common practice to use a transformation

of the price of a product as a proxy for its quality with the idea that a higher price would only be paid by customers if the product quality is also higher. To this end, consumer microdata is employed to deliver price data, often ignoring the fact that these datasets contain a high degree of sparsity.

Crucially, studies such as Kaplan and Menzio (2015a) and Pytka (2024) have shown that individual households often pay significantly different prices for identical goods. As highlighted by Burdett and Judd (1983), these price discrepancies arise due to price search frictions, as retailers weigh higher per-unit profits against increased sales volume. In such an environment—and given that only a few purchases may be observed for some products—the average price of these specific goods can be severely over- or underestimated, making it a poor proxy for product quality.

The primary goal of this project is to analyze whether search frictions in the consumption market, together with data sparsity, can lead to spurious quality differences attributed to products. I start with an illustrative example, after which I document some empirical patters supporting the idea that some of the observed quality differences are caused by random sampling and small sample bias. Building on this, I use a search-theoretical model to derive a theoretical pricing distribution which is then calibrated to fit real world consumption patterns. Using the calibrated distribution and the level of sparsity observed in real data, I can show that the presence of sparsity does indeed lead to spurious quality differences between products, caused by small sample bias for some of the very sparse products. In addition, systematic differences in price search behavior between high- and low-income households lead to a positive correlation between the quality proxy and the share of high-income households that consume a given product.

While the setting addressed in the simulation may seem relatively narrow at first glance, its implications are much broader. This is because similar issues with spurious results are to be expected in other contexts that are structurally similar and feature choice sets significantly larger than the observed number of entities making these choices, along with a significant degree of sparsity.

Literature Review. This project relates closely to three strands of literature. It is in line with a growing body of literature emphasizing the relevance of sparsity in large-scale datasets. Gentzkow et al. (2019) highlight the issue in the context of

measuring polarization in political speech, while Pytka ⓡ Runge (2025) address the issue in the context of polarization in consumption choices.

The retailer side of the model used to derive a distribution of posted prices is built in the spirit of Burdett and Judd (1983), while the consumer side is assumed to follow the structure of Pytka (2024). These papers are part of a growing literature highlighting the importance of price search frictions in macroeconomic applications.

In addition, the project is directly related to the literature measuring the quality of products utilizing price data. Some recent examples in this vein are Argente and Lee (2021) and Becker (2024). The quality of a product is measured as some transformation of the price of this product, and in most cases, prices of similar products are used to normalize quality and make it comparable across different classes of products.

This paper highlights the issues that the construction of quality measures based on transactional data might face when the dataset used features a high degree of sparsity. While I employ the Kilts-NielsenIQ Consumer Panel for calibration, the conclusions drawn are not exclusive but apply to a broader range of applications involving sparse datasets.

II. Illustrative Example

To illustrate the mechanism that leads to artificial results in the presence of data sparsity, I want to start by presenting a highly simplified example. Imagine that there are three products, which are identical in terms of utility and consumed by a group of consumers subject to price search frictions. The price search frictions cause price dispersion because some consumers only have one price offer to choose from, while others can decide between multiple offers. This incentivizes sellers to post different prices, trading off higher sales against higher per-unit profits.

Since all three products are identical, the same holds true for the distribution of prices posted by sellers for each of these products. Notably, since the three products are identical by assumption, the posted price distributions will be identical as well.

Imagine now that for each of these products, we observe a very limited number

of purchases and the corresponding prices. We can imagine these three products being part of a larger dataset, representing three of the products with a very high degree of sparsity. Figure 3.1 shows a possible set of drawn price points for each of the products, together with the data-generating process (DGP) from which these points are drawn.

If one were to infer the quality of the three products from the observed prices, one would conclude that Product 1 is the highest-quality product, Product 2 is of medium quality, and Product 3 is of low quality. Importantly, this finding would be completely artificial since the underlying posted price distributions are identical for all three products, and all differences are merely the result of random sampling.

The next section will present some empirical evidence supporting the idea that at least some of the quality differences are caused by random sampling and are not related to inherent differences between the products.

III. Empirical Patterns

In a first step, I want to describe the dataset I use within this project. One of the most used data sources for consumer microdata within the last years has been the the Kilts-NielsenIQ Consumer Panel (KNCP). The KNCP is an evolving panel of American households covering about 40,000 households in the years 2004-2006 and 60,000 from 2007 onward. They track each household's purchases either via in-home scanning devices or a mobile application and link each product purchase to a distinct shopping trip. In addition, NielsenIQ provides weights such that the panel is representative at the national level. Each product is identified by its unique barcode (UPC) and part of a distinct product module. There are over 2 million unique products grouped into about 1500 modules. Product modules collect similar products which are close substitutes. An example for a product module is `FRUIT JUICE - GRAPEFRUIT - FROZEN`, which collects all frozen grapefruit juices in the dataset.

If the product quality differences in the dataset documented within the literature are not caused by differences in true product quality, but instead are the result of the interaction between price dispersion and data sparsity, we should be able to find empirical patterns supporting this idea. If we assume the extreme case where

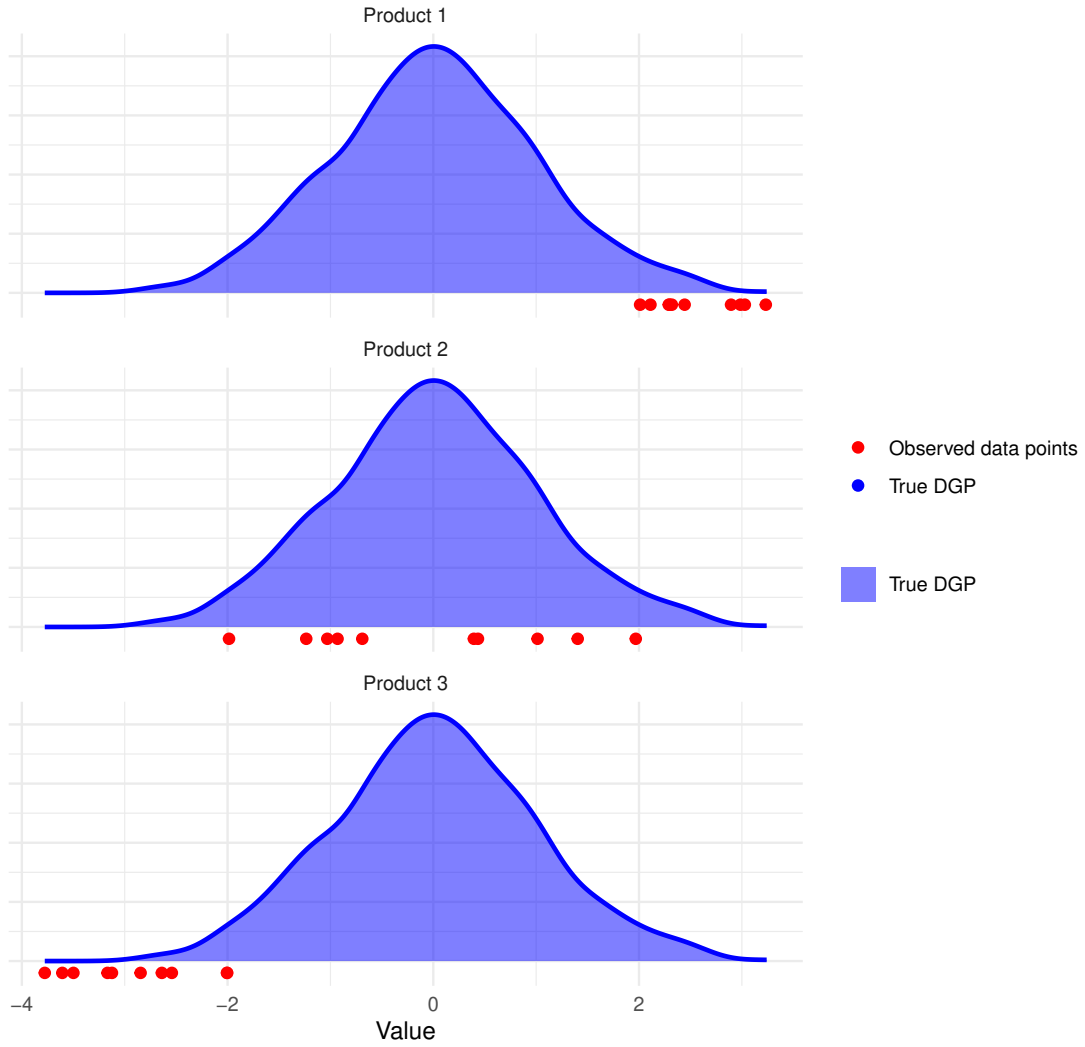


Figure 3.1: The figure shows how random sampling with a small sample size for three products with the same DGP can generate very different observed data points

there is no quality difference between products, then given a sufficiently large price sample, quality differences should disappear. This would suggest that there should also be a negative relationship between sample size and the variance of product quality.

In the following, we will check whether the data exhibits this negative relationship. Therefore, we first compute two different product quality measures. The first

measure follows the approach of Argente and Lee (2021). To formally define this product quality measure, let $p_{j,t}$ denote the price of product j at time t, and let $\hat{p}_{m,t}$ be the average price of product module m at time t, where m is the product module to which product j belongs. The quality of product j is then defined as:

$$Q_{j,t}^{AL} = log(p_{j,t}) - log(\hat{p}_{m,t})$$

The second measure is given by:

$$Q_{j,t}^{R} = \frac{p_{j,t}}{\hat{p}_{m,t}}$$

In both cases, the main idea is to measure quality by comparing the average price of a product to the average prices of similar products. If a product is expensive compared to similar products, it is considered to be of higher quality because people, on average, are willing to pay a higher price for it. In addition, quality differences become comparable by using the average within-module price to standardize individual product average prices. We compute these measures using three different time splits. The underlying reason for this is that it is not intuitively clear how long the time periods should be within which product quality is computed.

As the baseline, we use all available data over the entire horizon available for a product to compute one product quality observation per product. To ensure that our results are not driven by this approach, we additionally compute product quality at the yearly as well as quarterly level. The results for the baseline version are presented here, while the results for the yearly and quarterly levels are presented in Appendix A.

In the first step, we plot the densities for both quality measures to establish that there is indeed significant dispersion within quality measures. Furthermore, we see that the distribution is surprisingly symmetric, which is the shape that would be expected if all differences were caused by random sampling. This is especially true for the measure based on the log-differences. Note that the density of the ratio-based measure is truncated since there are some very high values that would make the plot unreadable.

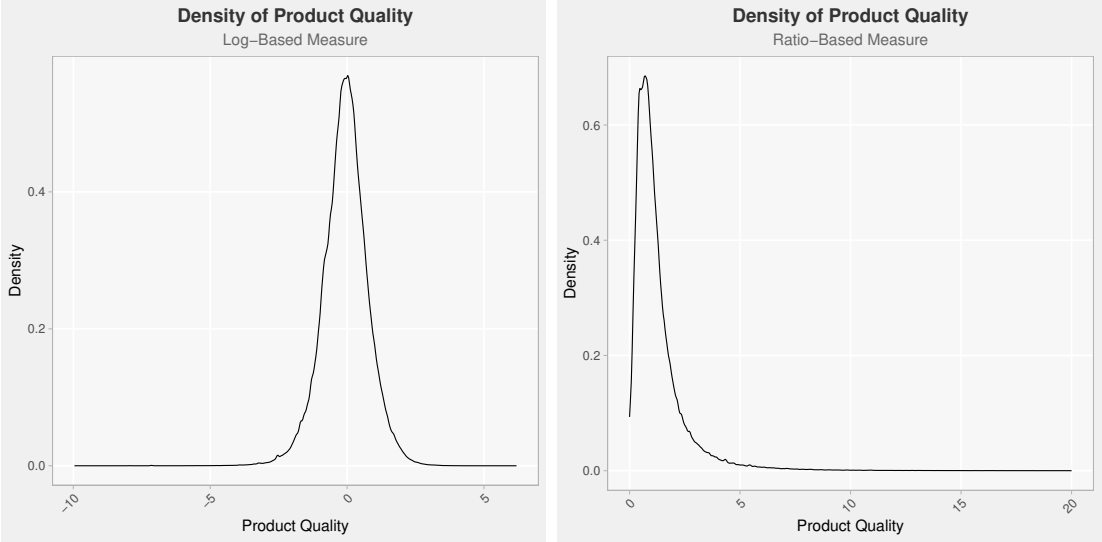


Figure 3.2: The plots display the density of both product quality measures. The plot for the ratio based measure is truncated at 20 to ensure readability.

From there, the next step is to compute measures of dispersion depending on the number of observed transactions for the different products. To measure dispersion depending on the number of shopping trips we first collect for each product the number of trips this product was purchased in. We then split the dataset into 10 subsets using the deciles of the number of shopping trips as the splitting criterion. For each of the subsets we then compute the tenth and ninetieth percentiles of the quality distribution, as well as the standard deviation. Both measures clearly show that dispersion declines significantly as the number of shopping trips increases, as can be seen in Figure 3.5. The corresponding standard deviations show a very similar pattern (see Appendix A). This provides clear evidence that at least some of the observed quality differences result from random sampling and are not caused by fundamental differences between the products.

IV. Price Search Model

To pursue this idea further, the next section will develop a stylized model to derive a paid price distribution, which we can then use jointly with the sparsity observed in the data to generate synthetic price observations under the assumption that all

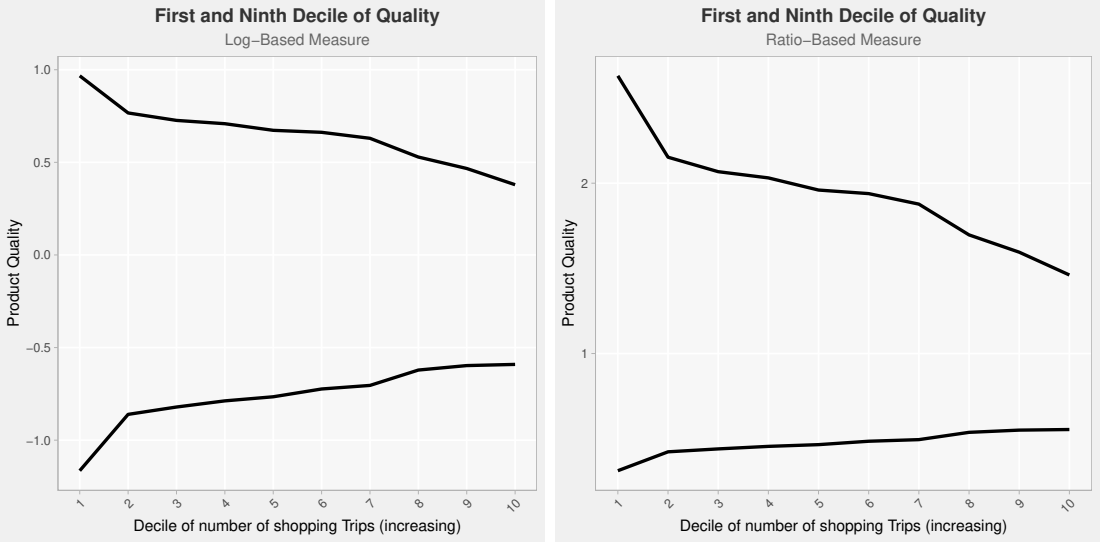


Figure 3.3: The plots show the tenth and ninetieth percentile of the product quality distribution for different deciles of the number of observed trips to purchase a product.

products are identical. Since we can use the sparsity exactly as it is observed in the data and match the pricing distribution to real-world data moments, this exercise allows us to further understand how much of the "quality" differences are due to product attributes and how much is the result of sparsity and random sampling.

The primary purpose of the model is to provide a foundation for the pricing distribution used within the simulation study, as well as to allow for the calibration of the distribution. For this purpose, a partial equilibrium model of the retailers' problem is sufficient, taking the choices of households as given.

Therefore, I abstract from explicitly solving the household side and assume that households solve a problem similar to Pytka (2024). There are a total of N different products in the economy, and each product is sold by a different group of retailers. All of these retailers are ex ante identical. The N products are identical in terms of utility, and each retailer sells only one specific product. Households are assumed to engage in price search as in Pytka (2024). With probability s, a household samples two price offers, and with the corresponding probability $1 - s$, they sample one price. If a household has multiple price offers, it will choose to accept the lower offer.

Retailers are modeled in the spirit of Burdett and Judd (1983). Within the market for each product, there is a continuum of retailers of mass 1. Since all markets are structurally identical, I present the retailer's problem here without indices referring to the different goods. Each retailer lives for one period and is able to procure the product at a constant marginal cost of 0.

When making the pricing decision, each retailer faces two different information asymmetries. The retailer is neither informed about the income of the household they meet nor whether the household has one or two offers to purchase the product. Each retailer is visited by households with equal probability and then offers the product to each visiting household at price p. The households can then choose to either buy the product at price p or forgo the purchase.

Denote by $G(p)$ the CDF of the price distribution, which is taken as given by each retailer. The price search intensity of households is denoted by s_h for the high-income type and s_l for the low-income type. Similarly, the amount purchased by high-income households is denoted by c_h, and c_l denotes the amount purchased by low-income households.

Let h be the fraction of high-income households, and correspondingly, $1-h$ the fraction of low-income households. Given this, profits are given by:

$$\pi = h\left(1 - \frac{2s_h}{1+s_h}G(p)\right)pc_h + (1-h)\left(1 - \frac{2s_l}{1+s_l}G(p)\right)pc_l$$

The structure of the retailer's problem is similar to that in Burdett and Judd (1983), expanded to two types of households with potentially different probabilities of sampling either one or two price offers and potentially different quantities purchased. As the following lemma shows, the results from Burdett and Judd (1983) can be extended to this setup.

Lemma 1 *The CDF of the price distribution G(p) exhibits the following properties:*

1. *G(p) is continuous.*
2. *profits are identical for all posted prices.*
3. *the highest price posted is equal to the household's reservation price.*

Proof. The proof can be found in Appendix B.

Since all product markets are independent of one another, it is sufficient to define the equilibrium for a single product market. The equilibrium definition is given by:

Equilibrium Definition *Given household choices s_h,s_l,c_h, and c_l, the distribution $G(p)$ is such that:*

1. *Retailers post prices that maximize profits given household choices*

Using the general properties of the CDF and the equilibrium definition, I can then derive a closed-form solution:

Theorem 1 *Given household shopping behavior, equilibrium price dispersion can be expressed in closed form:*

$$G(p) = \begin{cases} 0, & for\ p\ \leq \underline{p} \\ \frac{h\frac{2s_h}{1+s_h}\bar{p}c_h - h\bar{p}c_h - (1-h)\bar{p}c_l + (1-h)\frac{2s_l}{1+s_l}\bar{p}c_l + hpc_h + (1-h)pc_l}{h\frac{2s_h}{1+s_h}pc_h + (1-h)\frac{2s_l}{1+s_l}pc_l}, & for\ p\ \in [\underline{p}, \bar{p}] \\ 1, & for\ p\ >\ \bar{p} \end{cases}$$

where $\underline{p}$ is given by:

$$\underline{p} = \frac{hc_h\bar{p}\left(1 - \frac{2s_h}{1+s_h}\right) + (1-h)c_l\bar{p}\left(1 - \frac{2s_l}{1+s_l}\right)}{hc_h + (1-h)c_l}$$

Proof. The proof can be found in Appendix B.

V. Analysis of Simulated Price Data

Now that we have established a theoretical foundation for the posted price distribution, the next step is to use this to generate an artificial dataset. This will be done as follows. The underlying parameters for all goods are identical, meaning that the amounts purchased by both types of households, as well as their search intensities, are the same across all markets. This is consistent with the assumption that the goods are perfect substitutes with identical properties, yielding the same

utility to households and therefore being of the same quality. For each market, I then use the calibrated distribution of prices from the model to generate a dataset of paid prices for both household types. From these, I then sample subsets of different sizes representing the sparsity observed in the data. Following the approach taken in the literature, I then compute the log-based measure for product quality already employed within the empirical analysis.

Based on the simulated data, I conduct two types of analysis. First, I check the presence and extend of quality differences the product quality measures suggest for the simulated data. Since in this case there is by construction no difference between the goods in terms of quality, this exercise is designed to determine whether the quality ladder itself can be caused solely by the presence of price search and data sparsity.

In a second step, I additionally investigate whether the finding that higher-income households consume higher-quality goods can be artificially generated in this setup. To do so, I use a naive OLS estimation to check for a link between the share of high-income households that purchase a given good and the average quality of that good. A more detailed description of how this is implemented can be found in Appendix C.

Before running the simulations, I calibrate the paid price distribution to capture data moments that identify the four underlying parameters: the quantities purchased by both groups and the search intensities. I normalize the quantity purchased per trip to one for the low-income group and set the parameter for the high-income group to match the average ratio in expenditures per trip between the low- and high-income groups. I compute the ratio directly from the data and it is given by 1.02. The two search intensities are set to match the relative price difference between high- and low-income households, as well as the variance of the average price distribution. For the relative price difference I match 7% higher prices for the high-income group which is documented in Pytka (2024) and the variance of the average standardized price distribution I match is 19%, which is documented in Kaplan and Menzio (2015b). Finally, the ratio of low-income to high-income households h is set to match the ratio between the number of shopping trips for both types of households. The ratio of trips is 0.623. The model matches the target moments perfectly.

I directly obtain the degree of sparsity from the data. I do so by computing, for each good in the dataset, the number of shopping trips in which that good is purchased. In my simulation, I simulate as many goods as there are in the NielsenIQ dataset, and I directly use the number of shopping trips to determine how many observations are sampled for each good. Therefore, the sparsity in the simulation perfectly reflects the level of sparsity in the data.

To start, I plot the density of both quality measures for the simulated data. When comparing the resulting plots to those from the real data, it becomes clear that dispersion is much lower in the simulated data. This suggests that dispersion within the quality measures computed from the real data is at least partially due to genuine quality differences.

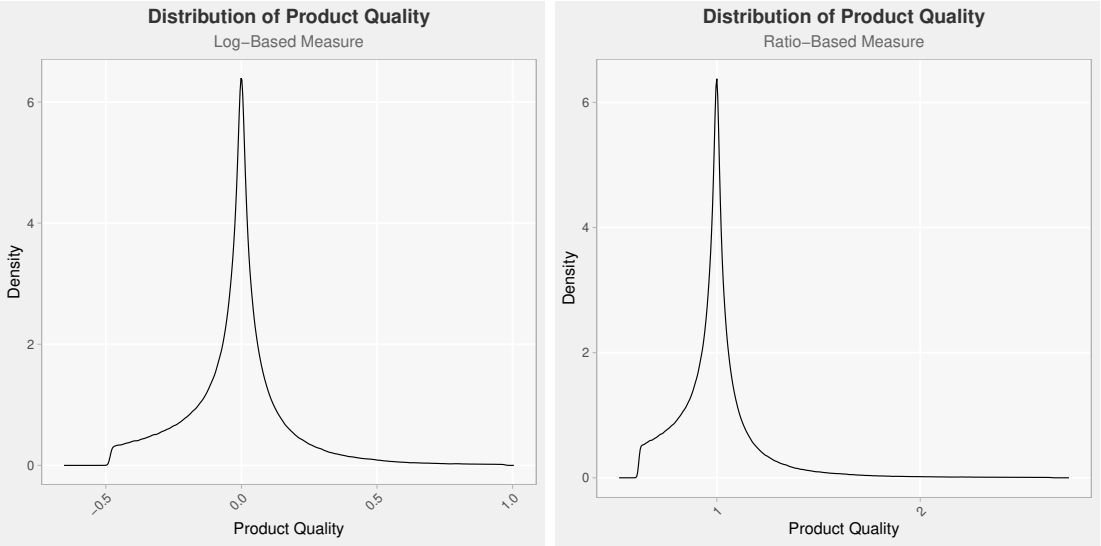


Figure 3.4: The plots display the density of both product quality measures for the simulated data.

To gain better insight into the degree to which quality differences are the result of sparsity and price search, I next examine the link between dispersion and purchasing frequency in the simulated data. The plots clearly show that dispersion drops much faster than in the real data. For the first decile, the simulated data generates about 20% of the dispersion for the log-based measure and about 30% for the ratio-based measure. As soon as the number of shopping trips increases beyond the first decile, dispersion in the simulated data almost disappears. This

clearly highlights that quality differences between products that are purchased very infrequently might be significantly overestimated.

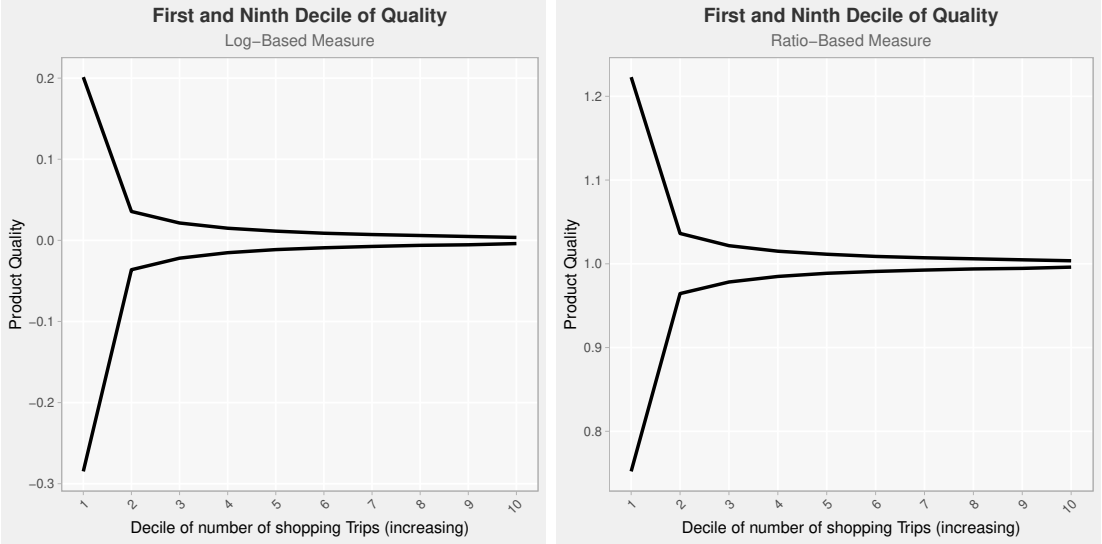


Figure 3.5: The plots show the tenth and ninetieth percentile of the product quality distribution for different deciles of the number of observed trips to purchase a product.

Finally, I use the simulated data to analyze whether the combination of data sparsity and the different price search intensities leads to a link between the fraction of high-income households that consume a product and the quality of this product. I do so by running a simple OLS regression explaining quality with the fraction of high-income households. If I find a positive coefficient for this relationship this indicates that higher "quality" is associated with a higher fraction of high-income households. This would suggest to the researcher that on average high-income households consume higher quality products then low-income households.

In both cases, the estimated coefficient is relatively similar, with a value of 0.004132 for the log-based measure and 0.004792 for the ratio-based measure. Both estimates are highly significant. This clearly shows that sparsity, jointly with the differences in price search, leads to dispersion in product quality as well as a spurious link between higher-quality goods and the fraction of high-income households, even though there is no difference between the products in utility terms.

VI. Concluding Thoughts

I documented that the distribution of product quality within the KNCP is surprisingly symmetric. Additionally, there is a clear negative relationship between dispersion in product quality and the number of shopping trips in which a product is purchased. This relationship hints at the fact that some of the dispersion in quality is caused by small sample bias, and as the sample size increases, estimates become more precise and therefore dispersion decreases.

I then used a calibrated model to generate synthetic data under the assumption that all goods are perfect substitutes. From this synthetic dataset, observations for individual products are sampled to accurately mimic the sparsity observed in the KNCP. Using this dataset, I show that the combination of data sparsity and price search frictions can lead to spurious quality differences between goods, even though the goods are identical in terms of utility. Notably, only for the most infrequently purchased goods is about one-fifth of the dispersion in quality differences observed in the real data artificially generated. For more frequently purchased goods, almost no artificial quality differences arise. This implies that most of the observed differences reflect genuine quality variation across goods. While this leaves significant room for a product quality ladder, the findings in Pytka ⓡ Runge (2025) suggest that no such ladder exists along the income dimension. Additionally, systematic price differences between households with different income levels can generate a spurious link between the fraction of high-income households observed consuming a given good and its estimated quality.

The key takeaway from this exercise is that, especially in datasets with significant degrees of sparsity, it is important to carefully select an appropriate estimation strategy to not overweight spurious quality differences. In estimation contexts, this could mean employing penalized estimation or some kind of weighting scheme that reduces the influence of products with very few observations.

Beyond these more broadly applicable implications, there is also one that is specific to the context of product quality and price search with differing search intensities. The finding that, even in the synthetic dataset, one would come to the conclusion that higher-income households consume higher-quality goods shows a clear flaw in the way quality is measured. The idea that a higher price signifies

higher quality because households are willing to spend more on these products is not wrong in itself. The problem in this context is that price search is also costly in terms of utility and potentially more costly when a higher volume of goods is consumed. Under these circumstances, price differences for identical goods will arise due to different types of households consuming these goods, even if they ultimately yield the same utility.

Bibliography

Argente, David and Munseob Lee (2021) "Cost of Living Inequality during the Great Recession," *Journal of the European Economic Association*, 19 (2), 913–952.

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

Burdett, Kenneth and Kenneth L Judd (1983) "Equilibrium Price Dispersion," *Econometrica*, 51 (4), 955–69.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy (2019) "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 87 (4), 1307–1340.

Kaplan, Greg and Guido Menzio (2015a) "The Morphology of Price Dispersion," *International Economic Review*, 56 (4), 1165–1206.

——— (2015b) "The Morphology of Price Dispersion," *International Economic Review*, 56 (4), 1165–1206.

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

Pytka, Krzysztof ⓡ Daniel Runge (2025) "Looking into the Consumption Black Box: Evidence from Scanner Data," Working paper.

3.A. Appendix

A. Robustness

This section contains additional material from the empirical section. It mainly contains the results for the two additional specifications where product quality is computed on a yearly as well as a quarterly basis. As the plots clearly show, the exact numbers change between the specifications but the main findings stay consistent. The densities remain relatively symmetric and the standard deviation within persistence declines in the number of observed transactions, as can also be seen by the first and ninth deciles.

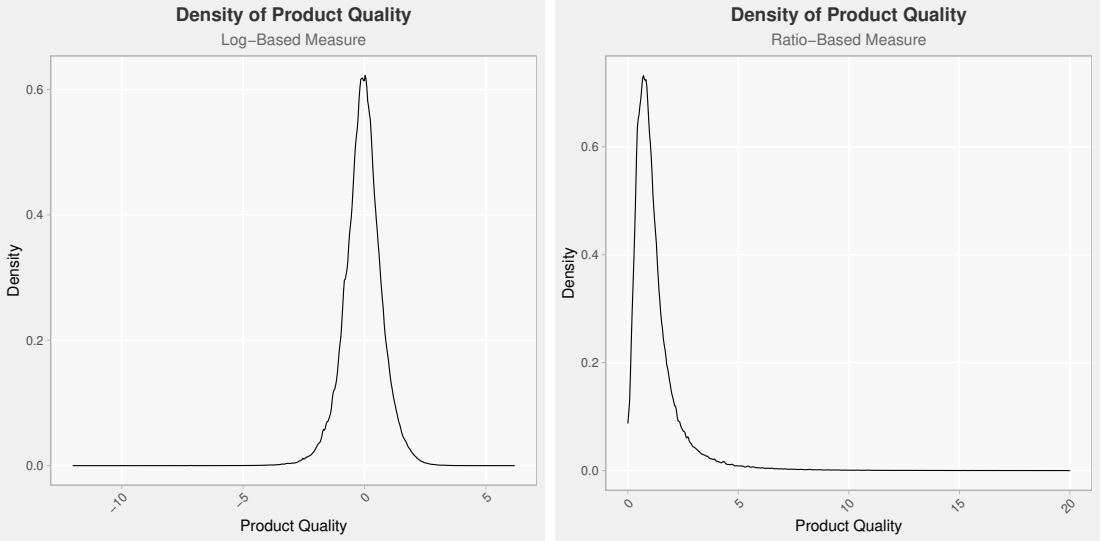


Figure 3.6: The plots display the density of both product quality measures. The plot for the ratio based measure is truncated at 20 to ensure readability.

Additionally, I provide the standard deviations for all the quality measures. The results stay consistent with the conclusions from analyzing the Tenth and Ninetieth percentile.

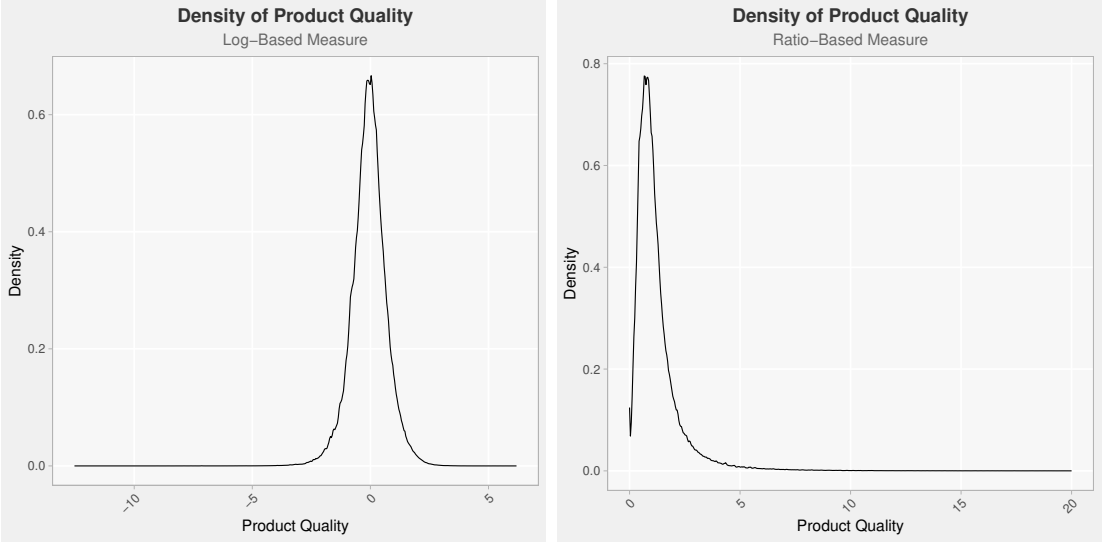


Figure 3.7: The plots display the density of both product quality measures. The plot for the ratio based measure is truncated at 20 to ensure readability.

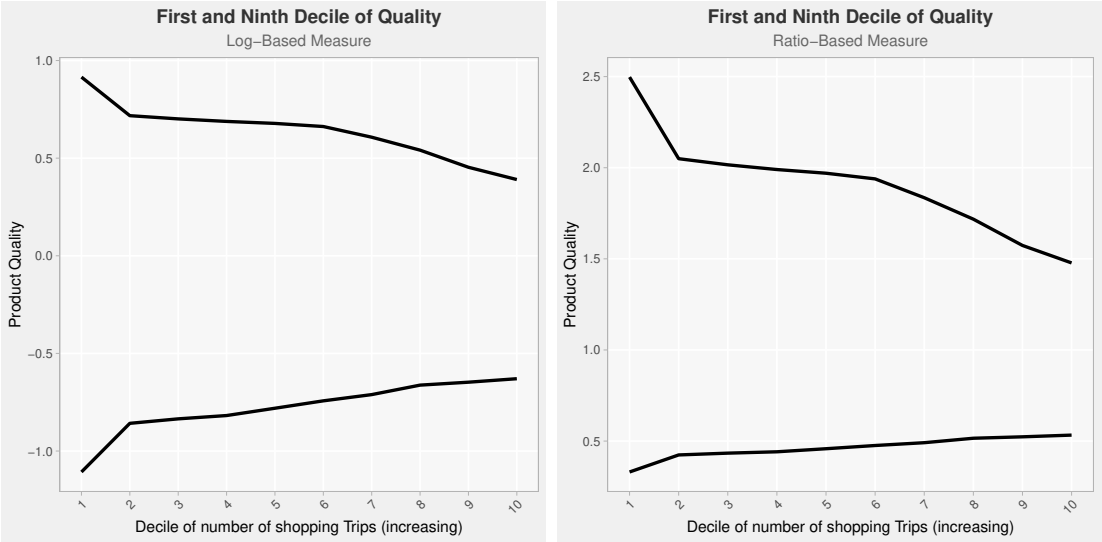


Figure 3.8: The plots show the tenth and ninetieth percentile of the product quality distribution for different deciles of the number of observed trips to purchase a product.

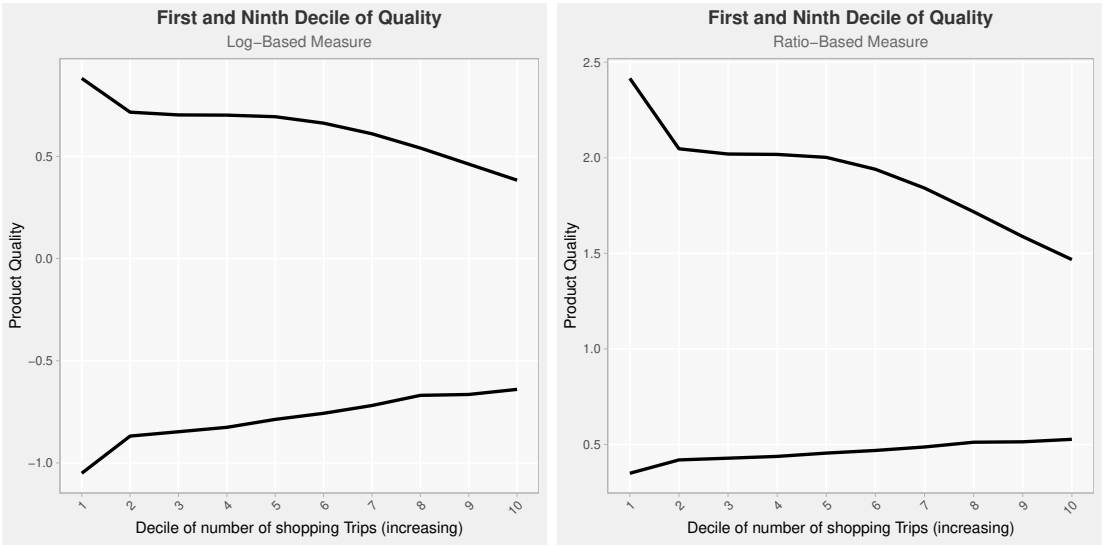


Figure 3.9: The plots show the tenth and ninetieth percentile of the product quality distribution for different deciles of the number of observed trips to purchase a product.

Frequency	10%	20%	30%	40%	50%
All Data	0.8702651	0.6721898	0.6384721	0.6177207	0.5963232
Yearly	0.8360726	0.6601868	0.6406677	0.6257016	0.6079391
Quarterly	0.7997569	0.6682369	0.6476984	0.6360310	0.6160571
Frequency	60%	70%	80%	90%	100%
All Data	0.5802052	0.5624286	0.5000244	0.4561039	0.4175087
Yearly	0.5915856	0.5532009	0.5152188	0.4728319	0.4289030
Quarterly	0.5964809	0.5594914	0.5187274	0.4816081	0.4363067

Table 3.1: Standard deviations for the log-based product quality measure.

Frequency	10%	20%	30%	40%	50%
All Data	1.7188327	1.0596151	0.9883793	0.9906596	0.9681244
Yearly	1.6438934	1.0291908	0.9865693	0.9597457	0.9621040
Quarterly	1.5211069	1.0369905	0.9960226	0.9790509	0.9895378
Frequency	60%	70%	80%	90%	100%
All Data	1.0026942	0.9594371	0.8106656	0.6869599	0.4384877
Yearly	1.0345588	0.9224565	0.8464741	0.5635369	0.4485926
Quarterly	1.0396240	0.9219227	0.8390735	0.5831971	0.4484234

Table 3.2: Standard deviations for the ratio-based product quality measure.

Proof of Lemma 1

1. G(p) is continuous.

This follows directly from Lemma 1 in Burdett and Judd (1983). Assume $G(p)$ is not continuous. Then, when a retailer reduces its price by an infinitesimal amount $p-\epsilon$, the probability of making a sale would change by a discrete amount, since $G(p)$ is discontinuous. A profitable deviation from the equilibrium strategy would therefore exist, which is a contradiction.

2. the profit is identical for all posted prices.

Assume this is not the case. Then, there exist prices p' and p'' such that $\pi(p') > \pi(p'')$. In this case, it would be profitable for a retailer posting p' to deviate to posting p''. This contradicts p' being posted in equilibrium. Therefore, no such pair of prices can exist in equilibrium, and it must be the case that profits are equal.

3. the highest price posted is equal to the households reservation price.

Assume the highest price posted, $\bar{p}$, is not equal to the household's reservation price p^r. For the highest price posted, we have that $G(\bar{p}) = 1$. There are two cases: either $\bar{p} < p^r$ or $\bar{p} > p^r$.

In the second case, profits from posting the price $\bar{p}$ would be zero, since the price is above the reservation price and, therefore, households do not make any purchases. Then, the retailer could deviate to any price below the reservation price, which would yield a positive profit. Thus, $\bar{p}$ being posted by a retailer is a contradiction to equilibrium conditions.

For the case $\bar{p} < p^r$, profits from posting $\bar{p}$ are equal to:

$$\pi(\bar{p}) = h(1 - \frac{2s_h}{1+s_h})\bar{p}c_h + (1-h)(1 - \frac{2s_l}{1+s_l})\bar{p}c_l$$

since the retailer only sells to captive consumers. The profits from instead posting the reservation price are given by:

$$\pi(p^r) = h(1 - \frac{2s_h}{1+s_h})p^r c_h + (1-h)(1 - \frac{2s_l}{1+s_l})p^r c_l$$

Since $\bar{p} < p^r$, it must be the case that profits from posting $\bar{p}$ are lower than profits from posting p^r, which is a contradiction. Therefore, the upper bound of the price distribution must be equal to the household reservation price.

Proof of Theorem 1

To derive the price distribution, I use Properties 2 and 3 established in Lemma 1. The profit from posting any price p must be equal to the profit from posting the upper bound of the price distribution, $\bar{p}$. This gives:

$$\begin{aligned} h(1 - \frac{2s_h}{1+s_h}G(p))pc_h + (1-h)(1 - \frac{2s_l}{1+s_l}G(p))pc_l \\ = h(1 - \frac{2s_h}{1+s_h})\bar{p}c_h + (1-h)(1 - \frac{2s_l}{1+s_l})\bar{p}c_l \end{aligned}$$

Solving this equation for the CDF $G(p)$ yields the result:

$$G(p) = \frac{h\frac{2s_h}{1+s_h}\bar{p}c_h - h\bar{p}c_h - (1-h)\bar{p}c_l + (1-h)\frac{2s_l}{1+s_l}\bar{p}c_l + hpc_h + (1-h)pc_l}{h\frac{2s_h}{1+s_h}pc_h + (1-h)\frac{2s_l}{1+s_l}pc_l}$$

This proves the first part of the theorem. For the lower bound of the price distribution, I solve:

$$\frac{h\bar{p}c_h - h\frac{2s_h}{1+s_h}\bar{p}c_h + (1-h)\bar{p}c_l - (1-h)\frac{2s_l}{1+s_l}\bar{p}c_l - h\underline{p}c_h - (1-h)\underline{p}c_l}{-h\frac{2s_h}{1+s_h}\underline{p}c_h - (1-h)\frac{2s_l}{1+s_l}\underline{p}c_l} = 0$$

This yields:

$$\underline{p} = \frac{hc_h\bar{p}\left(1 - \frac{2s_h}{1+s_h}\right) + (1-h)c_l\bar{p}\left(1 - \frac{2s_l}{1+s_l}\right)}{hc_h + (1-h)c_l}$$

which completes the Theorem.

C. Simulation Algorithm

In this part, I will describe the simulation in more detail. In a first step, I calibrate the parameters as described in the main text. I then simulate as many products as there are unique products in the dataset. For each of these products, I do the following:

1. I generate about 250,000 price observations. Each observation is generated by:
   (a) Drawing two uniform random numbers between 0 and 1
   (b) Using the inverse CDF of the posted price distribution to find the corresponding price
   (c) Draw an additional random number that is uniform between 0 and 1
   (d) If that random number is less than the search intensity, I use the minimum of the two prices, otherwise the first one sampled
2. From these price observations, I then randomly sample the prices for the final dataset. The number of observations sampled is equal to the number of observed shopping trips for this good from the original dataset.
3. I repeat this procedure until all 250,000 goods are simulated.

Daniel Runge

Lebenslauf

Ausbildung

Seit 2019	**Promotion in Volkswirtschaftslehre**, *Universität Mannheim*
2017–2019	**M.Sc. in Volkswirtschaftslehre**, *Universität Mannheim*
2013–2017	**B.A. in Staatswissenschaften (Volkswirtschaft, Sozialwissenschaften und Recht)**, *Universität Erfurt*
2010–2013	**Abitur**, *Privates Johannes Gymnasium Lahnstein*

Berufserfahrung

Seit 09.2015	**PhD Trainee**, *Europäische Zentralbank*
2020–2024	**Teaching Assistant**, *Universität Mannheim*
2016–2017	**Teaching Assistant**, *Universität Erfurt*