Inequality in Income and Varieties

Ha Bui*
The University of Texas at Austin

Job Market Paper

This version: October 20, 2023. The latest version here.

Abstract. This paper studies the impact of local income inequality on household consumption and welfare through the changing availability of varieties of consumer goods. I find that in high-inequality counties, more varieties are offered to households. However, households living in high-inequality counties purchase fewer varieties compared to similar ones in low-inequality counties. This happens even though households shop in more stores in high-inequality counties. These effects are even more pronounced for individuals in the tails of the income distribution. To quantify channels underlying these empirical findings and speak to the welfare impact, I develop a model featuring an endogenous number of varieties produced by firms and a choice over which varieties to purchase by households. After estimating many of the parameters of the model, I show that the model can reproduce both empirical facts. Quantitatively, I find that households are generally worse off when living in higher-inequality regions because firms are better able to segment the market.

Keywords: Income Inequality, consumption inequality, variety, market segmentation.
JEL Codes: D12, D31, D63, E21

*habuithu@utexas.edu. I am indebted to my advisors, Olivier Coibion, Nitya Pandalai-Nayar, and Christoph Boehm, for their continuous guidance and support. This work benefited greatly from comments by Daniel Ackerberg, Saroj Bhattarai, Corina Boar, Kristy Buzard, Andres Drenik, Nathasha Egodage, Stefano Eusepi, Keith Head, Cooper Howes, Kevin Kuruc, James Lake, Andrei Levchenko, Julian Ludwig, Andreas Mueller, Brent Neiman, Dimitrije Ruzic, Aysegul Sahin, Anastasia Zervou, and conference and workshop participants at UT-Austin, the Southern Economic Association Annual Meeting, and Midwest Macro Meetings. I thank the Texas Advanced Computing Center (TACC) at UT-Austin for providing HPC resources. All errors are mine. Researcher owns 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.

1 Introduction

Income inequality in the U.S. has been rising over the past fifty years (e.g., Chancel and Piketty, 2021).[1] There is a vast literature studying whether consumption inequality has gone up along with income inequality (e.g., Krueger and Perri, 2006; Heathcote et al., 2010; Aguiar and Bils, 2015; Attanasio and Pistaferri, 2014, 2016; Meyer and Sullivan, 2023). This literature mostly focuses on the intensive margin of household consumption which is measured as how much households spend on consumption.

Yet the literature has overlooked another fundamental aspect of consumption: the extensive margin, that is, the set of varieties consumed by households. Consumption should be considered in two aspects: (1) what to consume – the varieties household consumes, and (2) how much to consume – the spending. The difference in the set of varieties available to consume would have different implications for household welfare. Even with the same level of spending, household welfare could increase if there are more products available to consume as predicted by love of variety (e.g., Dixit and Stiglitz, 1977; Krugman, 1979, 1980; Helpman and Krugman, 1985), or households could choose to consume products that are closer to their ideal varieties and achieve higher welfare as in the ideal-variety model(e.g., Lancaster, 1979, 1980; Hummels and Klenow, 2005).

This paper studies how local income inequality affects household consumption and welfare through the changing availability of grocery products. I aim to answer three central questions: (1) do high-inequality regions have more varieties than low-inequality ones?; (2) do households living in high-inequality regions purchase more varieties compared to similar ones living in low-inequality areas?; and (3) what are the implications of the answers to (1) and (2) for household welfare?

In answering these questions, the paper makes three main contributions. Empirically, using detailed matched home and store scanner consumption microdata, I document two novel facts. First, more varieties of products are sold in high-inequality counties than in lower inequality counties. Second, at the individual level, a household living in a high-inequality county tends to

[1] Figure 1 plots the Gini Index in the U.S. over the period 1967-2020 using US Census Bureau data and we can see a clear upward trend indicating rising income inequality.

consumer fewer varieties than would a comparable household living in a low-inequality county.[2] Theoretically, I develop a model with an endogenous number of varieties produced by firms and a choice over which varieties to purchase by households to quantify the channels underlying these puzzles and speak to the welfare impact. I then use the microdata to estimate model parameters, most of which have not been estimated in the literature, and show that the model can replicate both empirical facts. Quantitatively, using the estimated parameters, I quantify the impacts of variety inequality on household welfare and find that households are generally worse off when living in higher-inequality areas because firms are better able to segment the market.

The analysis hinges on a comprehensive geographic dataset from NielsenIQ over the period of 2010-2020. This dataset includes the Consumer Panel data, a nationally representative survey of grocery purchases of about 60,000 households, and the Retail Scanner data containing detailed sales information from 30,000 to 50,000 stores, accounting for approximately 40 percent of total U.S. grocery sales. The richness in regional heterogeneity of these data ensures the cross-sectional variation needed to identify inequality in varieties. Moreover, the dataset also has a well-defined definition of variety as product barcode or brand, which allows us to trace the set of products available in a county and the set of products purchased by households.

Using the NielsenIQ data, I document two puzzling empirical facts showing the relationship between inequality in income and varieties: high-inequality counties have more varieties, but households in these counties purchase fewer varieties. First, there are more varieties available in high-inequality counties than in low-inequality counties. It is true even at more disaggregated levels such as the number of varieties by category or by store, and at more aggregated levels such as the number of categories or the number of stores. Second, despite having more varieties available, households living in high-inequality counties consume fewer varieties than similar ones living in low-inequality areas. The results hold at more disaggregated levels such as varieties by category or by store, and at more aggregated levels such as the number of categories. However, although households in high-inequality counties purchase fewer varieties, they tend to frequent a greater number of stores. These effects are even more pronounced for households in the tails of the income

[2] The baseline geographic unit of the analysis is county as it is easy for households to shop at different zip codes. Another reason for choosing county in the baseline analysis is because we need reliable location-specific variables data which other geographic units like commuting zone or city could not satisfy this data requirement. However, in Appendix B, I perform robustness check at the commuting zone level and the results are qualitatively similar.

distribution. The same patterns hold when we focus on people who move from one county to another.

One channel that could explain these puzzling facts is the “ideal variety channel" which is the concept that households concentrate their spending on their favorite varieties instead of trying different varieties available in the market. For counties with the same level of population and other characteristics, retailers in high-inequality counties face demands from consumers with considerably more income heterogeneity than in low-inequality counties. Thus, retailers will tend to offer a broader spectrum of products to accommodate demand. Having access to more varieties makes it easier for households in high-inequality counties to find products closer to their ideal varieties, so they concentrate their spending on their favorite goods instead of trying various products. In this case, household welfare improves when living in high-inequality counties.

Another channel that could rationalize these puzzling facts is the “market segmentation channel" which is the concept that the market in high-inequality counties is more segmented. For counties with the same level of population and other characteristics, high-inequality counties would have fewer households with a similar level of income. If firms provide products for their targeted group of customers, then for a household of a certain level of income, living in a high-inequality county means there are fewer products targeted to their group even though there are more varieties in total. If households love to consume varieties targeted at them, they will consume fewer varieties as there are fewer suitable options for them in a high-inequality county. In this case, households will tend to be worse off when living in high-inequality counties.

The empirical findings show that both channels exist in the data. On one hand, the fact that households in high-inequality counties shop at more stores even though they purchase fewer varieties suggests the existence of the “ideal variety channel". Households in high-inequality counties shop at different stores to get products closer to their ideal varieties. On the other hand, we have another piece of evidence supporting the “market segmentation channel". We observe that the effects of variety inequality are more pronounced for households in the tails of the income distribution. It suggests that there is differential treatment for different groups of households. As inequality goes up, more goods are being targeted for the two ends of the income distribution, so we see bigger changes for them than for the middle group of the income distribution.

Since both channels coexist but they have different implications for household welfare, I develop a model with an endogenous choice of variety by households and firms to quantify the impact of variety inequality on household welfare. On the demand side, the model incorporates income inequality by assuming household income is drawn from a log-normal distribution with the same mean but different variances. I specify non-homothetic preferences over grocery products and allow households across the income distribution to differ in their evaluations of product quality attributes. Inspired by Li (2021) and Neiman and Vavra (2023), I allow for a disutility term that represents the cost of consuming varieties so that households would only choose to consume products whose quality is above a certain threshold of household's taste parameter. This allows us to derive an equation that determines the optimal number of varieties consumed by a household, which depends on the number of available varieties. On the supply side, following Melitz (2003) and Faber and Fally (2022) I allow firms to be heterogeneous in their productivity, and thus, in the quality of the product they produce. Both marginal and fixed costs can be functions of output quality, allowing for economies of scale in production. Firms now operate in a setting where their pricing and quality choices affect the composition of market demand that they face. This allows us to derive an equation showing the relationship between the number of firms in a county and the county's income inequality level. The model admits closed-form solutions which enables us to explain the observed data moments and quantify the impact on household welfare.

Next, I use the NielsenIQ data to estimate the model parameters many of which have not been estimated in the literature yet. *First*, the estimated parameters show that the richer place higher value on product quality attributes. Also, producing higher quality product entails higher cost. These conditions incentivize firms to segment the market and only produce higher-quality products if the mass of targeted customers is high enough to justify the higher cost. *Second*, the estimated parameter of household's elasticity of substitution reveals that households have inelastic substitution preferences, indicating that they find products less substitutable. Moreover, the cost of consuming varieties increases in the number of varieties consumed, so household would only include a product in their consumption basket if the utility gain exceeds this cost. These estimated parameters support the existence of the "ideal variety channel". *Thirdly*, the cost of consuming varieties increases with the number of varieties consumed and the number of varieties available not targeted for household type, but decreases in in the number of available varieties targeted for household type. These estimated parameters support the existence of the "market segmentation channel".

Finally, armed with these estimates I quantify the impact of inequality in varieties on household welfare. I find that the "market segmentation channel" dominates the "ideal-variety channel". Consequently, households experience a decline in their welfare when residing in higher-inequality areas. Overall, my findings suggest that having access to more varieties does not necessarily mean household welfare would improve as the predictions in the love-of-variety model.

My work contributes to a large existing literature studying whether the increase in income inequality was mirrored by consumption inequality (e.g., Krueger and Perri, 2006; Heathcote et al., 2010; Aguiar and Bils, 2015; Attanasio and Pistaferri, 2014, 2016; Coibion et al., 2015, 2021; Meyer and Sullivan, 2023). All of these studies focus on the intensive margin of consumption and employ household expenditure data to measure consumption inequality. Different from them, my research concentrates on the extensive margin of consumption, that is, the set of products consumed by households. My contribution is to explore the relationship between income inequality and this extensive margin of consumption, and shows that rising income inequality affects the set of products purchased by households.

My results contribute to the existing literature on the relationship between regional heterogeneity and variety availability. Previous research has established that the number of products available in a region is influenced by factors such as population size (e.g., Hummels and Klenow, 2005; Bernard et al., 2007; Handbury and Weinstein, 2015), local income level (e.g., Alwitt and Donley, 1997; Horowitz et al., 2004; Algert et al., 2006; Karpyn et al., 2019; Allcott et al., 2019; Handbury, 2021; Smets et al., 2022), and the presence of multi-region retailers (e.g., Gilbert, 2017; Hyun and Kim, 2019). I extend this line of work and study the association between variety availability and income inequality within a region. I utilize NielsenIQ Retail Scanner Data tracking grocery store sales, which allows for accurate calculations of variety availability. After controlling for variables such as county population, median household income, and other location-specific characteristics, my analysis shows evidence suggesting that high-inequality counties have more varieties.

This paper also relates to a growing literature studying the heterogeneity of varieties across different income groups (e.g., Kaplan and Schulhofer-Wohl, 2017; Jaravel, 2018; Argente and Lee, 2021; Faber and Fally, 2022) and within income brackets (e.g., Li, 2021; Neiman and Vavra, 2023). In contrast to those studies, my paper emphasizes the differences in varieties within an income group but living in dissimilar regions. My contribution to this line of research is to provide empirical evidence that the local income distribution is associated with variety heterogeneity even within an income group.

Finally, this work complements the literature in industrial organization which develops models to study the impacts of income inequality on market outcomes in vertically differentiated markets. My finding that high-income inequality counties have more varieties than low-income inequality counties is consistent with model prediction in Gabszewicz and Thisse (1979, 1980), Shaked and Sutton (1982), Benassi et al. (2006), and Yurko (2011) who show that higher inequality in consumer incomes leads to the entry of more firms and more intense competition among the entrants. My paper also relates to the empirical literature in industrial organization showing the availability of varieties depends on the preference of the majority in the market (Waldfogel, 2003; George and Waldfogel, 2003; Waldfogel, 2007). My results are about the availability of varieties depending on consumers' characteristics, in particular, their income distribution.

The paper is organized as follows. Section 2 describes the NielsenIQ data and county-level data. Section 3 presents two empirical findings showing the effects of income inequality on the number of varieties available in counties and the number of varieties purchased by households. Section 4 develops a theoretical model to rationalize the empirical findings. Section 5 presents parameter estimation. Section 6 analyzes welfare implications. Section 7 concludes.

2 Data

To study the relation between income inequality and variety inequality, I use NielsenIQ data which tracks weekly store sales and consumer purchases in 49 US states. This enables us to identify specific products sold by stores and purchased by households at detail levels of barcode and brand. I also incorporate county-level characteristics from American Community Survey (ACS) and Bureau of Labor Statistics (BLS) to control for other factors in the regressions.

2.1 NielsenIQ Data

The analysis in this paper use NielsenIQ Retail Scanner and Consumer Panel Data[3] in 2010-2020. The Retail Scanner Data (RMS) consists of weekly pricing and volume sold at 30,000-50,000 stores from approximately 90 retail chains across all U.S. markets. The Consumer Panel Data (HMS) tracks a panel of approximately 60,000 households in over 2,700 counties in 49 US states, except Alaska and Hawaii, with their demographic characteristics and their purchases of consumer goods from a wide range of retail outlets. Appendix A contains a detailed description of the datasets.

I restrict the study to all years in the period of 2010-2020. The Consumer Panel Data is available from 2004. The Retail Scanner Data is available from 2006. I use data from 2010 onward as the analysis requires information from both datasets and some location-specific variables[4] which are only available from 2010 from ACS and BLS. The geographic unit of analysis is the county as it is easy for households to shop at stores located in different zip codes, whereas shopping across different counties can be challenging.

These NielsenIQ datasets have three useful features: a well-defined definition of variety, detailed information on household demographic and geographic characteristics, and a wide coverage of the US market. First, they have a well-defined notion of product variety as a Universal Product Code (UPC) or a brand. UPCs are defined by manufacturers with unique identified product features such as packaging, size, color, and flavor. University of Chicago - Kilts Center assigns a brand code for each UPC. Information on prices and quantities sold by store and purchased by households for each UPC is also available, which allows us to calculate product expenditure share. Second, HMS contains household demographic and geographic information[5], so I can map household characteristics to products to study different household consumption patterns along the income distribution. Third, these datasets have the richness in regional heterogeneity, which ensures cross-sectional variation in income inequality and product differentiation to identify the relationship of interest. The RMS data covers more than half of the total sales volume of U.S.

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

[4]The location-specific variables used are population, unemployment rate, poverty rate, median household income, density, and the percentage of black, white, and Hispanic population at the county level.

[5]The HMS data has information on household characteristics including household income brackets, household composition, household size, presence of children, information on household members, age, occupation, education, race, and Hispanic origin of the head of household. It also contains information on zip code, county code, and state where the household resides.

grocery as well as drug stores and more than 30 percent of all U.S. mass merchandiser sales volume. The HMS data provides sampling weights to make it demographically representative of the broader U.S. population.

Definition of product variety. This paper defines a product as a UPC in the baseline results. The results are robust when defining a product as a brand. The measure of product varieties is the number of UPCs or brands counting along different dimensions of product hierarchy. This definition benefits from simplicity while capturing the idea of measuring the scope of products offered by retailers and consumed by households. I exclude a subset of the data that does not use standard UPC codes such as fruits, vegetables, meats, in-store baked goods, and other random weight items.[6]

Product hierarchy. Figure 4 shows the product hierarchy in NielsenIQ data: departments, product groups, and product modules. The smallest level of aggregation above UPC or brand in this paper is product module while the most aggregated level is department. For example, "Haagen-Dazs Strawberries Cream Ice Cream 460ml" is a UPC within the product module "Ice cream - bulk" within the product group "Ice cream, novelties" within the department "Frozen foods".

2.2 County-level characteristic data

Besides the barcode-level NielsenIQ data, I employ county-level data from the ACS 5-year estimates and BLS to account for income inequality and other location-specific variables in our regression analyses.

Income inequality. The paper uses the county-level Gini index from the ACS 5-year estimates as the measure of local income inequality in the baseline regressions. The 5-year estimates include data collected over a 60-month period. The date of the data is the end of the 5-year period. For example, a value dated 2014 represents data from 2010 to 2014. However, they do not describe any specific day, month, or year within that period.

I focus on local income inequality measures for similar reasons as explained in Coibion et al. (2020). First, a local measure at the county level avoids measurement issues when comparing

[6]NielsenIQ calls this subset of data magnet data and provides separate sample weights for it. The sample weights used in the analyses in this paper are for UPC-coded data.

income across different areas. Second, income is likely the most common measure of well-being when households compare themselves to others. Third, the variation in income inequality across counties is essential to isolate the effects of inequality on household consumption behaviors.

I use the 5-year estimates because they are published for counties with populations of all sizes and are the most reliable and precise of the ACS period estimates. The 1-year estimates provide data on counties with populations of 65,000+. Hence, the 1-year estimates only have data for approximately 840 counties out of more than 3,100 counties in the US and miss more than 74% of counties in the US. The 3-year estimates provide data on counties with populations of 20,000+ but they were discontinued with the 2011-2013 release.

Although the Gini coefficient could be calculated using the HMS data, I use the Gini index from ACS data as it is a more precise measure of local income inequality. The HMS data provides information on household income by range. I could use the income bins and the fraction of households in each bin to calculate the Gini coefficient. However, the HMS data only covers a fraction of the U.S. population, and the income ranges are measured with a two-year lag relative to the observed shopping transactions. Thus, it would not be as updated and precise as the Gini index from ACS.

For robustness check, I also use the ratio of the mean income for the highest quintile (top 20 percent) of earners divided by the mean income of the lowest quintile (bottom 20 percent) of earners in a county as the measure of local income inequality. The mean income by quintile data is also from ACS 5-year estimates. I use the Gini index as the baseline measure instead of this measure because the interpretation for the Gini index is more straightforward.

Other location-specific variables. Besides the Gini index, the regressions also control for a set of county-level variables. They are population, median household income, poverty rate, unemployment rate, density, rural-urban dummy variable, and the percentage of black, white, and Hispanic population. Unemployment rate data are from the Bureau of Labor Statistics. Median household income and poverty rate data are from the U.S. Census Bureau's Small Area Income and Poverty Estimates program and ACS 5-year estimates.[7] County density is calculated by dividing population by land area in square miles. Land area data is from US Census data. The rural-urban

[7]All the results in Section 3 are robust when I control for mean household income or income per capita instead of median household income. The income per capita data is from the U.S. Bureau of Economic Analysis. The mean household income data is from ACS 5-year estimates.

classification data is from the National Center for Health Statistics (NCHS). Data for other variables are from ACS 5-year estimates.[8]

I control for these county-level variables to limit the influence of confounding factors to the relationship between income inequality and variety. In particular, I control for population and density following the lead from Handbury and Weinstein (2015) who show that there are more products in cities with a larger population. I control for the unemployment rate and poverty rate following Malmendier and Shen (2018) who show that households who have lived through times of high local and national unemployment, or who have experienced more personal unemployment, spend significantly less on food and total consumption. Including median household income as an independent variable helps us to focus on the effect of income inequality – the only difference across counties is the variance of the income distribution while we hold the median fixed. Finally, I control for rural-urban dummy variable, and the percentage of black, white, and Hispanic population following Desmet and Wacziarg (2022) as the set of varieties offered in a region is affected by the composition of its population.

Table 1 contains summary statistics of the Retail Scanner Data for the number of varieties and the number of stores available in counties by quartiles of some county-level variables. The average numbers of varieties and stores increase as the Gini coefficient, population, and median household income at the county level go up while they show a decreasing trend when the unemployment rate and poverty rate rise. The summary statistics of the Consumer Panel Data for the number of varieties consumed by low- and high-income households by quartiles of location variables are found in Table 2. High-income households, on average, consume more varieties than low-income households. When the Gini index, population, and median household income at the county level increase, the average number of varieties purchased by both low- and high-income households reduces.

3 Empirical Findings: Variety Inequality

Using the NielsenIQ data, I conduct reduced-form analyses to study the relationship between income inequality and variety inequality. The results reveal a surprising pattern: high-inequality

[8] Appendix A contains a detailed description of the data.

counties have more varieties, but households in these counties consume fewer varieties compared to similar households living in low-inequality counties.

3.1 Income Inequality and Varieties in Counties

The first analysis examines the impact of income inequality on product availability in counties. I aggregate weekly Retail Scanner Data into yearly data and conduct the analysis at both the county and store levels.

Varieties by county. The baseline specification estimating the relationship between the number of products available in a county and its income inequality level is as follows.

$$log(\Sigma_{rt}) = \alpha + \beta \text{Gini}_{rt} + \gamma_c C_{rt} + \text{State-Year FE} + \varepsilon_{rt} \tag{3.1}$$

where Σ_{rt} is the number of product barcodes (UPCs) sold by all stores in county r in year t. Gini_{rt} is the Gini index of county r in year t, which measures the level of income inequality in the county. The parameter of interest is β, which describes the relationship between the number of varieties available in a county and its level of income inequality. The vector of location controls, C_{rt}, includes population, median household income, unemployment rate, poverty rate, density, rural-urban (metro) dummy variable, and the percentage of black, white, and Hispanic population in county r in year t. State-Year FE refers to state-year fixed effects, which capture changes in the state over time. The standard errors are clustered by county. The results are reported in Table 3 - Panel A. The first finding is that the estimated coefficient on the Gini index is positive, which implies that holding other variables constant, high-inequality counties have more varieties. In particular, if the Gini coefficient increases by 0.1, the number of varieties available in a county on average increases by 50% while holding everything else fixed.

The result that the more-unequal counties have more varieties could be attributed to both the extensive margin — the increase in the number of categories or stores, and the intensive margin — the increase in the number of products offered within a category or a store.

Extensive and intensive margins by category. First, I analyze the effects due to the extensive and intensive margins along the category dimension. To distinguish between these two effects, I use a similar estimation as in the specification (3.1). However, for the extensive margin analysis, the dependent variable is the number of categories offered in a county, and for the intensive margin analysis, it is the number of UPCs within a category in a county. Additionally, I control for category-state-year fixed effects in the latter regression. Categories can be defined as either departments, product groups, or product modules. Table 3 - Panel B presents the results, which show that the increase in the number of varieties in a county is attributed to both the extensive and intensive margins. Counties with higher income inequality have more categories and more varieties within a category. The interpretation of the estimated coefficients, for example, when the category is defined as a product group, is that when the Gini index goes up by 0.1, the number of product groups in a county increases by almost 1.5% and the number of varieties within a product group in a county increases by roughly 57% on average, while holding everything else fixed.

Extensive and intensive margins by store. The higher number of varieties available in high-inequality counties could also be attributed to the extensive margin — the increase in the number of stores, or the intensive margin — the increase in the number of varieties offered within each store in the area. To disentangle these effects, I employ a similar estimation as in the specification (3.1). However, for the extensive margin analysis, the dependent variable is the number of stores in a county, while for the intensive margin analysis, it is the number of UPCs available in a store. In the regression with the dependent variable being the number of UPCs available in a store, I also control for the fixed effects of the type of retailer, where the retailer type is defined as one of the six "channel codes" in the NielsenIQ data: mass merchandisers, drug stores, food stores, liquor stores, convenience stores, and convenience stores that have gas stations. Table 3 - Panel C presents the results. Once again the estimated parameters β are positive and statistically significant. They indicate that there are more stores in high-inequality counties and those stores also offer a greater variety of products compared to stores in low-inequality regions. On average, when the Gini index increases by 0.1, the number of stores in a county increases by almost 70% and the number of varieties in a store increases by 5% while holding everything else fixed.

Extensive and intensive margins by store by category. We can look further into the increase in the number of varieties offered in a store as the Gini index goes up. It can be because the stores in high-inequality counties offer more categories or offer more products within each category. Using the same regression as above, we obtain the results in Panel D - Table 3. They show that stores in high-inequality counties have more categories and carry more varieties within each category. The interpretation of the estimated coefficients, taking the example when the category is defined as a product group, is that when the Gini index goes up by 0.1, while holding everything else constant, the number of product groups offered by a store increases by 1.4% and the number of varieties within a product group in a store increases by nearly 6% on average.

I also conduct a robustness check of all the analyses above using brand as the definition of variety. The results are reported in Table 11 in Appendix B. We find similar results that high-inequality counties, on average, have more brands of product, more brands within a category, more brands in a store, and more brands within a category in a store.

These findings establish the first empirical result: high-inequality counties have a broader selection of varieties than low-inequality ones, and it is driven by both the extensive and intensive margins across different dimensions.

3.2 Income Inequality and Varieties purchased by households

Having documented that the number of varieties available in a county rises with income inequality, I now study how the number of varieties purchased by households varies with income inequality.

3.2.1 Income Inequality and Purchased Varieties

I conduct the analyses for the sample of all households, poor households, middle-income households, and rich households. I perform the analyses for different income groups because it might be possible that the effects of variety inequality is different for different income groups.

In the baseline estimate, poor households are defined as those with income in the bottom 20^{th} percentile of the national income distribution in the dataset each year, rich households are those with income in the top 80^{th} percentile of the national income distribution, and the rest is the middle-income group. In particular, the poor are those with an income of at most $\$24,999$, the rich are those with an income of at least $\$100,000$, and the middle group are those with an income of at least $\$25,000$ to below $\$100,000$. This definition ensures that the threshold of classification by income is the same for households in different regions. I also conduct a robustness check using the thresholds of income groups defined with the regional income distribution and the results are similar. All the results are also robust when we change the thresholds in this definition to 30% and 70%, or 10% and 90% for the poor and the rich, respectively.

Varieties consumed by household. The baseline regression showing how the number of varieties purchased by households varies with the level of income inequality of counties is as below.

$$log(\Sigma_{rit}) = \alpha + \beta \text{Gini}_{rt} + \gamma_h X_{rit} + \gamma_c C_{rt} + \text{State-Year FE} + \varepsilon_{rit} \tag{3.2}$$

where Σ_{rit} denotes the number of varieties purchased by household i living in county r in year t. The vector of household characteristics, X_{rit}, includes household income, household size, type of residence, gender, age, employment, education, occupation, marital status, race, Hispanic origin of the head of the household, the presence of children under age 18, and the number of children under age 18. C_{rt} is the vector of location-specific controls, including population, median household income, unemployment rate, poverty rate, density, rural-urban dummy variable, and the percentage of black, white, and Hispanic population of county r where household i lives in year t. Other variables are the same as in previous specifications. The parameter of interest is β, which captures the relationship between the number of varieties consumed by households and the levels of income inequality of counties. The standard errors are clustered by county and household. The results are presented in Table 4 - Panel A. The estimated coefficients of the Gini index are negative for all households and different income groups. It implies that households in high-inequality counties, on average, purchase *fewer* varieties compared to similar households living in low-inequality counties holding other variables fixed. In particular, as the Gini ratio increases by 0.1, the number of varieties consumed by households living in high-inequality regions decreases by approximately 5% compared to similar ones residing in low-inequality regions.

Another way to examine this question is to nest the first result into it by testing if households purchase more varieties when there are more products available. To this end, I estimate another specification replacing the Gini index with the log of the number of varieties offered in a county (Σ_{rt}). To address the concern that it could be possible for households to shop at neighboring counties, I control for the log of the average number of varieties available in neighboring counties of county r in year t ($\bar{\Sigma}_{rt}$). Other variables are the same as in the regression (3.2).

$$log(\Sigma_{rit}) = \alpha + \beta_1 log(\Sigma_{rt}) + \beta_2 log(\bar{\Sigma}_{rt}) + \gamma_h X_{rit} + \gamma_c C_{rt} + \text{State-Year FE} + \varepsilon_{rit} \tag{3.3}$$

The results of this estimation retain the same sign as the results of specification (3.2) and are reported in the Table 5. The estimated coefficients of Σ_{rt} are interpreted as the number of varieties consumed by households decreases when the number of varieties available in a county increases, holding everything else fixed.

This finding contradicts the prediction of the standard love-of-variety model. To better understand what drives this result, I disentangle the effects by analyzing how the lower number of varieties consumed by households in high-inequality counties is driven by the change in extensive and intensive margins along different dimensions.

Extensive and intensive margins by category. First, the result that households in high-inequality counties consume fewer varieties could be because they consume fewer categories or fewer products within a category. To disentangle these two effects, I use similar estimations as in the specification (3.2). However, the dependent variable in the extensive margin analysis is the number of categories consumed by a household, and the dependent variable in the intensive margin analysis is the number of UPCs within a category purchased by a household. I also control for category fixed effects in the intensive margin analysis. The results in Table 4 - Panel B show that both the extensive and intensive margins are at work. On average, households living in high-inequality counties purchase fewer categories and fewer varieties within a category when holding other variables constant. The interpretation of the estimated coefficient, for example when category is defined as product group, is that when holding everything else fixed, if the Gini index goes up by 0.1, the number of product groups consumed by a household decreases by $2 - 3.6\%$; and the number of varieties in a product group consumed by a household decreases by roughly $2.6 - 7\%$ on average.

Extensive and intensive margins by store. The fact that households in more-unequal counties consume fewer varieties could also be due to the extensive margin — they shop at fewer stores, or the intensive margin — they buy fewer products from each store. I employ similar estimation as in specification (3.2) to disentangle these effects. The dependent variable is the number of stores that a household shops at for the extensive margin analysis, and the number of UPCs a household buys in a store for intensive margin analysis. In the latter regression, I also control for retailer fixed effects. Table 4 - Panel C presents the results of these estimations. Interestingly, households in high-inequality counties shop at more stores, but they buy fewer products per store than similar households living in low-inequality areas. In particular, holding everything else fixed, on average, when the Gini index increases by 0.1, households shop at $3.5 - 9\%$ more stores, but they purchase around $8 - 12\%$ fewer products per store.

Extensive and intensive margins by store by category. The fact that households purchase fewer products per store could be because they purchase fewer categories or because they buy fewer products in a category. Using the same regression as above, we obtain the results in Panel D - Table 4. It shows that households consume fewer products by store because of both the extensive and intensive margins. The interpretation of the estimated coefficient, for example when category is defined as product group, is that when the Gini index goes up by 0.1, holding everything else fixed, the number of product groups purchased by a household decreases by around 5% and the number of varieties in a product group consumed by a household decreases by nearly $2 - 6\%$ on average.

All the results above hold when we conduct the analyses for different income groups, as shown in Table 4. I also conduct robustness check of all the analyses above using brand as the definition of variety. The results are reported in Table 14 in Appendix B. We find similar results that households in high-inequality counties, on average, consume fewer brands, fewer brands within a category, fewer brands in a store, and fewer brands within a category in a store.

Together these results establish the second empirical fact that households living in high-inequality counties purchase *fewer* varieties compared to similar households in low-inequality regions, and it is driven by both the extensive and intensive margins across different dimensions. However, they do so while shopping at more stores.

3.2.2 Income Inequality and Movers' Purchased Varieties

One might be concerned that a household with a certain level of income and other household characteristics living in a high-inequality county would not be the same as another household of the same level of income and other characteristics residing in a low-inequality county. To address this concern, I look into movers who move from one county to another. Since households move for different reasons which might cause endogeneity concerns, I further restrict the sample to movers whose income change is only within the range of $\pm\$10,000$, and there is no change in their marital status and the number of children. There are 7,236 households in this sample. I then employ the following regression to study the effect of moving to a higher income inequality counties on the number of varieties households purchase.

$$\Delta log(\Sigma_{rit}) = \alpha + \beta_1 \Delta \text{Gini}_{rt} \times \text{Post move} + \beta_2 \Delta \text{Gini}_{rt} + \beta_3 \text{Post move} + \gamma_h \Delta X_{rit} + \gamma_c \Delta C_{rt} + \varepsilon_{rit} \quad (3.4)$$

where $\Delta log(\Sigma_{rit})$ is the difference in the log of number of varieties purchased by household i in year t compared to the log of number of varieties purchased by this household in the base year. Similarly, ΔGini_{rt} is the change in the average value of the Gini index of the county across all years where household i lives before the move compared to the average value of the Gini index of the county across all years where household i lives after the move. Post move is a dummy variable which has the value of 1 if it is the year after household i moves, and the value of 0 if it is the year before household i moves. ΔX_{rit} captures any changes in the characteristics of household i before and after the move (for example, small changes in household income). Finally, ΔC_{rt} captures other differences between the county where household i lives before the move and the new county where household i lives after the move. The coefficient we are interested in is β_1,which captures how the number of varieties purchased by households changes when moving to a county with higher level of income inequality. The standard errors are clustered by household and county.

I perform the analysis using different base years, i.e, year 3 before the move, year 4 before the move, and year 5 before the move. The negative coefficient of β_1 suggests that as households move to counties with higher level of income inequality, they purchase fewer varieties. In particular, if the Gini index of the destination county is higher than the Gini index of the original county by 0.1, the number of varieties bought by households decreases by around 9%.

The analysis for movers provides another approach to assess the relationship between the level of income inequality of counties and the number of varieties bought by households. The results further support our second empirical finding that households living in high-inequality counties purchase *fewer* varieties.

3.3 Discussion

So far we have two key empirical findings. There are more varieties available in high-inequality counties. However, households living in high-inequality counties purchase fewer varieties. These results are at odd with the prediction of the standard love-of-variety model which predicts that consumers would consume more varieties when there are more products available because they have diminishing marginal utility for each variety. I find that the empirical facts suggest that there are two plausible channels rationalize these two puzzling facts.

Ideal-variety channel. The first channel that could explain these puzzling facts is based on the idea of "ideal variety" consumption. It is the concept that households prefer to concentrate their consumption on their favorite products instead of trying different products available in the market. For counties with the same level of population, retailers in high-inequality counties face demands from consumers with a wider range of income heterogeneity than in low-inequality counties. Thus, retailers would offer a broader spectrum of products to accommodate demand. Having access to more varieties makes it easier for households living in high-inequality counties to find products closer to their ideal varieties, so they concentrate their spending on their favorite varieties instead of trying different products. Therefore, this channel suggests that living in high-inequality counties will improve household welfare because they could consume products closer to their ideal varieties.

The finding in the sub-section 3.2.1 - Extensive and intensive margins by store - that households in high-inequality counties shop at more stores suggests the existence of the ideal-variety channel. Households in high-inequality counties have access to more stores, so households now go to different stores to get products closer to their ideal varieties.

Market segmentation channel. The second channel that could rationalize these puzzling facts is based on the ideal of market segmentation. It is the concept that households love to consume more varieties which are targeted for them, so the more targeted varieties are available to consume, the greater their well-being is. For counties with the same level of population, high-inequality counties would have fewer households with a similar level of income. If retailers know their market well, they will offer more targeted products to each specific group of households when there are more households of this type in the county. Thus, in high-inequality counties, the markets are more segmented. For a household of a certain level of income, living in a high-inequality county means they have access to fewer products targeted for their type. If households prefer to consume varieties that are targeted toward them, they will consume fewer varieties as there are fewer options suitable for them in a high-inequality county. This channel suggests that households are worse off when living in high-inequality counties because the market is more segmented and there are fewer product targeted for them.

In Table 4, we observe that the effects are more pronounced for rich and poor households than for the middle-income group. I test whether there is significant difference in the estimated coefficients between middle-income households and rich households or between middle-income households and poor households. I can strongly reject the hypothesis that the effects are the same for these income groups. I also test if the difference in the estimated coefficients between the rich and the poor is statistically significant. The results of p-value do not reject the hypothesis that the effects are the same for the rich and the poor. In short, these results suggest the existence of the market segmentation channel.

Another way to test for the existence of the market segmentation channel is to see whether households would consume more varieties if more varieties targeted for them are available. To implement this exercise, I assume that the rich tend to consume high-quality products, and the poor tend to consume low-quality products, following the empirical finding in Handbury (2021). Then I employ the following regression to study the relationship between the number of varieties households purchase and the number of varieties targeted for the households:

$$log(\Sigma_{rit}) = \alpha + \beta_{1j} log(\Sigma_{rjt}) + \beta_{2j} log(\bar{\Sigma}_{rjt}) + \beta_{3j} log(\Sigma_{r\_jt}) + \beta_{4j} log(\bar{\Sigma}_{r\_jt}) + \ + \gamma_h X_{rit} + \gamma_c C_{rt} + \text{State-Year FE} + \varepsilon_{rit} \quad (3.5)$$

where Σ_{rit} is the number of UPCs purchased by household i living in county r in year t. Σ_{rjt} for $j = \{P, R\}$ is the number of low-quality (Σ_{rPt}) or high-quality (Σ_{rRt}) UPCs available in county r in year t, and Σ_{r_jt} is the number of high-quality (Σ_{rRt}) or high-quality (Σ_{rPt}) UPCs available in county r in year t. To address the concern that households might shop in neighboring counties, I control for $\bar{\Sigma}_{rjt}$ and $\bar{\Sigma}_{r\_jt}$ for $j = \{P, R\}$ which are the average number of low-quality (P) and high-quality (R) UPCs available in neighboring counties of county r in year t. Other variables are the same as previous specifications. Table 6 presents the results of this regression for the rich and the poor. The positive coefficients show that the more varieties targeted for households are available, the more varieties households purchase. On average, if the number of varieties targeted for households increases by 1%, the number of varieties purchased by households increases by about $0.012 - 0.019\%$.

These findings above show that both of the channels exist in the data, and they have different implications for household welfare. If the ideal variety channel dominates the market segmentation channel, household welfare will improve when moving to higher income inequality counties, and vice versa. In order to speak to the welfare impact, we need to develop a model which incorporates both channels.

4 A Model Linking Income Inequality and Variety Inequality

To rationalize the channels underlying the empirical facts found in the section 3 and quantify the welfare impact, this section develops a model featuring an endogenous number of varieties produced by firms and a choice over which varieties to purchase by households. On the demand side, following Li (2021) and Neiman and Vavra (2023), I allow for heterogeneous households who differ in their product quality evaluations and have non-homothetic preferences over grocery products. On the producer side, following Melitz (2003) and Faber and Fally (2022), firms are heterogeneous in productivities, face the observed distribution of consumer preferences and income, and optimally choose their product quality and prices. The model incorporates both channels: ideal-variety channel and market segmentation channel. Appendix C provides additional details of the model.

4.1 Household problem

I denote county as r. Counties have different levels of income inequality which is the Gini index. It is modeled as the difference in the variances of household income distributions, but they have the same mean. Each county has a continuum of household $i \in [0, 1]$ and has N_r varieties which is decided endogenously in the model.

The economy consists of two sectors: grocery sector and non-grocery sector.[9] Each household i living in the county r spends their income E_{ri} on the two sectors. As in Handbury (2021) and Faber and Fally (2022), I consider a two-tier utility where the upper-tier depends on utility from grocery and non-grocery consumption (C_{ri0}). Following the literature, I do not explicitly model this upper-level expenditure allocation decision for the sake of exposition, but I assume that the non-grocery good is normal. Household income is drawn from a log-normal distribution $E_{ri} \sim \text{log-normal}(\mu, \delta^2)$. There are two types of households $j = \{P, R\}$, where P denotes the poor and R denotes the rich. Among N_r varieties available in a county, there are N_{rj} of them are products targeted for household type j and the rest N_{r_j} are products less targeted for household type j. This is endogeneously decided in the model.

Households spend their grocery expenditure ($E_{ri} - C_{ri0}$) on a set of varieties $\Omega_{ri} \subseteq N_r$ with each variety is denoted as ω. The utility function of household i of type $j \in \{P, R\}$ living in the county r from grocery consumption is then defined by:

$$\begin{aligned} \max U_{ri\in j} &= \left(\int_{\omega \in \Omega_{ri}} \left(\varphi_{i\omega} c_{ri\omega} \right)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} - \underbrace{F \times (|\Omega_{ri}|)^{\varepsilon} N_{rj}^{\varepsilon_j(j)} N_{r\_j}^{\varepsilon_{\_j}(j)}}_{\text{Cost of consuming grocery varieties}} \\ \text{s.t.} \; &\underbrace{\int_{\omega \in \Omega_{ri}} c_{ri\omega} p_\omega d\omega}_{\text{Grocery expenditure}} = \underbrace{E_{ri}}_{\text{Income}} - \underbrace{C_{ri0}}_{\text{Non-grocery expenditure}} \end{aligned} \tag{4.1}$$

where household i's taste parameter for variety ω depends on quality of variety $\varphi_{i\omega} = {\phi_\omega}^{\gamma_j}$ with γ_j is quality evaluation of type j. The parameter σ is the elasticity of substitution across products. The term $F \times (|\Omega_{ri}|)^{\varepsilon} N_{rj}^{\varepsilon_j(j)} N_{r\_j}^{\varepsilon_{\_j}(j)}$ is the cost of consuming grocery varieties. F is a fixed cost term, the cost of consuming grocery varieties grows exponentially in the measure of varieties consumed $|\Omega_{ri}|$. This cost also increases if there are more varieties less targeted for household type j (N_{r_j}). However, this cost decreases if there are more varieties targeted for household type j (N_{rj}). The intuition here is that, if there are more varieties targeted for household type j, it will make it easier for household type j to obtain and consume these varieties, and thus, the cost of consuming them will go down. Similarly, if there are more varieties less targeted for household type j, it will make it harder for household type j to find products they prefer, and thus, the cost goes up.

[9]The non-grocery sector could be thought of as anything else that household consume besides grocery products. For instance, education, housing, other services, and so on.

Each household i needs to make two decisions. Firstly, a household chooses the optimal number of grocery varieties to consume $|\Omega_{ri}|$. Secondly, a household optimally allocates their expenditure for grocery across Ω_{ri} varieties to maximize their utility from grocery consumption. The household solves this two-decision problem backward.

Household's second decision. By solving the first order condition w.r.t. $c_{ri\omega}$, we obtain the optimal consumption choice of a household for each variety ω, which is similar to the solution of the standard CES-utility maximization problem:

$$c_{ri\omega} = \frac{(E_{ri} - C_{ri0})\, \varphi_{j\omega}^{\sigma-1}}{p_{\omega}^{\sigma} P_{ri}^{1-\sigma}} \tag{4.2}$$

where P_{ri} is the price index:

$$P_{ri} = \left[\int_{\omega \in \Omega_{ri}} \left(\frac{\varphi_{i\omega}}{p_{\omega}} \right)^{\sigma-1} d\omega \right]^{\frac{1}{1-\sigma}} \tag{4.3}$$

with p_{ω} is the price of product ω. Comparing two goods ω and $\omega^{'}$, relative expenditures of household are then given by:

$$\frac{p_{\omega} \cdot c_{ri\omega}}{p_{\omega'} \cdot c_{ri\omega'}} = \left(\frac{p_{\omega}}{p_{\omega'}} \right)^{1-\sigma} \left(\frac{\varphi_{i\omega}}{\varphi_{i\omega'}} \right)^{\sigma-1} \tag{4.4}$$

Household's first decision. Using the properties of the CES price index, we can rewrite the household's problem as:

$$\max_{|\Omega_{ri}|} \left(\frac{E_{ri} - C_{ri0}}{P_{ri}} \right) - F \times (|\Omega_{ri}|)^{\varepsilon} \, N_{rj}^{\varepsilon_j(j)} N_{r\_j}^{\varepsilon\_j(j)} \tag{4.5}$$

In order to solve this problem and obtain closed-form solutions of the number of grocery varieties household consumes, we need to express the price index as a function of $|\Omega_{r,i}|$. I assume that the household's price-adjusted tastes are distributed Pareto following Li (2021) and Neiman and Vavra (2023). In particular, $\tilde{\varphi}_{i\omega} = \frac{\varphi_{i\omega}}{p_{\omega}} \sim \text{Pareto}(\varphi_m, \theta)$ where $\varphi_m > 0$ is the minimum possible value of $\varphi_{i\omega}$ and $\theta > 0$ governs the shape of the distribution. The larger θ means a flatter distribution of price-adjusted taste.

$$\Pr\left(\tilde{\varphi}_{i\omega} < y\right) = G(y) = \begin{cases} 1 - \left(\frac{\varphi_m}{y}\right)^{\theta} & \text{if } y \geq \varphi_m \\ 0 & \text{if } y < \varphi_m \end{cases} \tag{4.6}$$

Following Pareto's principle,[10] we have:

$$\frac{|\Omega_{ri}|}{N_r} = \left[\frac{\varphi_m}{\tilde{\varphi}^*_{ri}}\right]^{\theta} \tag{4.7}$$

where $\tilde{\varphi}^*_{ri}$ denotes the cutoff of price-adjusted taste of varieties that household i living in region r decides to consume, that is, household i would only consume varieties with the minimum price-adjusted taste parameter of $\tilde{\varphi}^*_{ri}$. Substituting expression (4.7) and using the Pareto distribution assumption for price-adjusted taste parameter, we can rewrite the price index for household i as follows:

$$P_{ri} = \frac{1}{\varphi_m}\left(1 + \frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}} \left(|\Omega_{ri}|\right)^{\frac{1}{1-\sigma}} \left(\frac{|\Omega_{ri}|}{N_r}\right)^{\frac{1}{\theta}} \tag{4.8}$$

Now we can substitute the price index into the household's maximization problem and solve for the optimal number of varieties that household i would consume. The solution of this problem is:

$$|\Omega_{ri}| = \left[\underbrace{(E_{ri} - C_{ri0})}_{\text{Income Effect}} \underbrace{\left(1 - \frac{\sigma-1}{\theta}\right)^{\frac{1}{1-\sigma}} \varphi_m}_{\text{Average Price}^{-1}} \underbrace{\frac{1}{\varepsilon F}\left(\frac{1}{\sigma-1} - \frac{1}{\theta}\right)}_{\text{Cost Effect}} \underbrace{N_{r\_j}^{-\varepsilon_{\_j}(j)} N_{rj}^{-\varepsilon_j(j)}}_{\substack{\text{Market}\\ \text{Segmentation}\\ \text{Effect}}} \underbrace{N_r{}^{\frac{1}{\theta}}}_{\substack{\text{Variety}\\ \text{Effect}}}\right]^{\frac{1}{z}} \tag{4.9}$$

where $z = \varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$. This expression has five terms, each has an intuitive interpretation. I refer to the first term as the income effect. We observe in the data that the rich tend to consume more varieties than the poor. Modelling the grocery expenditure as a proportion of household income, this effect captures the observed pattern in the data. The second term is the inverse of

[10]Vilfredo Pareto (1848-1923) proposed that the number of people (N_x) with incomes higher than x can be modeled as a power law: $N_x = A/x^{\alpha} = Ax^{-\alpha}$. Let the total population be N_0 and let the minimum income be x_0. Then $N_0 = Ax_0^{-\alpha}$. Thus, we can write it in a proportion terms: $N_x/N_0 = (x_0/x)^{\alpha}$. I apply this principle where N_r is the total number of varieties available (equivalent to N_0 in the original equation) and $|\Omega_{r,i}|$ is the number of varieties consumed by household i (equivalent to N_x in the original equation).

average price. It summarizes the full distribution of price-adjusted tastes of products as if there were a single price for one unit of the full bundle. It says that the higher the average price is, the fewer varieties a household would purchase. The third term captures the effect of the cost of consuming varieties. If it is costly for households to get products, they will buy fewer varieties. The fourth term represents the effect of market segmentation. If there are more varieties targeted for the household, they will consume more varieties; while if there are more varieties targeted for other types of households, they will consume fewer varieties. Finally, the interaction between the market segmentation effect and the variety effect captures the ideal variety channel which we will analyze further in the subsection 4.3.

4.2 Firm problem

Following Melitz (2003) and Faber and Fally (2022), I assume that firms draw their productivity a from a cumulative distribution $G(a)$ upon paying a sunk entry cost $f > 0$. Firms are heterogeneous with different productivities. Each firm produces one variety ω under monopolistic competition. Thus, the number of firms in a county in the equilibrium equals the number of varieties available.

Each firm makes at most three decisions. First, firm pay the entry cost f and discover their productivity a. Second, upon entry with a productivity draw, a firm may decide to immediately exit and not produce. If a firm decides to produce, they need to decide the optimal level of product quality ϕ_ω they would produce based on the distribution of consumer preferences and income. Third, firm chooses their optimal price p_ω. We will solve firm's problem backward.

If the firm decides to produce, it will incur the following costs associated with the production of a quantity $C_{r\omega}$ with quality ϕ and productivity a. First, a variable cost increases with the quality of product and decreases with the firm's productivity $vc_\omega(\phi, a) = \frac{\phi^\xi}{a}$. The parameter $\xi \geq 0$ capture the elasticity of the cost increase with the level of product quality. Second, a fixed cost depends on quality $fc_\omega(\phi) = b\beta\phi^{\frac{1}{\beta}}$.[11] I adopt this simple log-linear parameterization for fixed cost following Hallak and Sivadasan (2013), Kugler and Verhoogen (2012), and Faber and Fally (2022) for tractability. Third, there is a fixed cost for production which does not depend on quality $fc_{0\omega}$.

Firm's third decision: choosing the optimal price. The demand that a firm faces is the total consumption of all households whose price-adjusted taste for this variety is higher than their cutoff

[11]Example of this type of fixed cost is cost associated with R&D or design.

threshold in the county. The firm's profit is as follows.

$$\Pi_{\omega} = \underbrace{p_{\omega} C_{r\omega}}_{\substack{\text{Revenue from both} \\ \text{types of HHs} \\ \text{who consume } \omega}} - \underbrace{vc_{\omega}}_{\substack{\text{Variable cost} \\ \equiv \frac{\phi^{\xi}}{a}}} C_{r\omega} - \underbrace{fc_{\omega}(\phi_{\omega})}_{\substack{\text{Fixed cost depending} \\ \text{on } \phi \text{ (ex: R \& D)} \\ \equiv b\beta\phi^{\frac{1}{\beta}}}} - \underbrace{fc_{0\omega}}_{\substack{\text{Fixed cost not} \\ \text{depending on } \phi \\ >0}} - \underbrace{f}_{\substack{\text{Entry cost} \\ >0}} \tag{4.10}$$

In the equilibrium, firm's optimal price is $p_{\omega} = \frac{\sigma}{\sigma-1} \cdot \frac{\phi^{\xi}}{a}$ where firm's markup is the standard markup under monopolistic competition environment $\frac{\sigma}{\sigma-1}$ with $\sigma > 1$.

Firm's second decision: choosing the optimal quality. Solving the firm's profit maximization problem with respect to their quality ϕ_{ω}, the optimal quality satisfies the following implicit equation:

$$\phi_{\omega} = \left[\frac{1}{b} \cdot \frac{\sigma - 1}{\sigma}(\tilde{\gamma} - \xi)R_{r\omega}\right]^{\beta} \tag{4.11}$$

where $R_{r\omega}$ denotes total sales of firm ω for both types of households. $\tilde{\gamma} = \frac{\gamma_P R_{rP\omega} + \gamma_R R_{rR\omega}}{R_{r\omega}}$ is the weighted average quality valuation of firm with productivity a.

The equation (4.11) shows that there are several forces that determine the optimal quality of product. First, when $\beta > 0$, the term $R_{r\omega}^{\beta}$ implies that larger sales induce higher optimal quality, which is the scale effect. Second, a firm's optimal quality depends on how much the firm's customer base value quality, which is captured by $\tilde{\gamma}$. The more a firm expects to sell to customer with high-quality valuation, the higher quality the firm would like to produce. If a firm instead expects to sell mostly to customer with lower quality valuation, firm would choose a lower product quality. Third, the optimal quality also depends on the cost structure. A higher elasticity of marginal cost to quality, ξ, induces lower optimal quality.

Firm's first decision: entry decision. A firm will enter the market as long as they expect to earn positive profit. Thus, the last firm who would enter the market is the one with zero profit. The equilibrium number of varieties in a county is characterized by a mass of N_r of firms (and hence, N_r varieties) and a distribution $h(a)$ of productivity levels over a subset of $(0, \infty)$.

In such an equilibrium, household's price index is as follows.

$$P_{ri\in j} = \left[\int_{\omega\in\Omega_{ri}} \left[\frac{\varphi_{i\omega}}{p_\omega}\right]^{\sigma-1} d\omega\right]^{\frac{1}{1-\sigma}} = \left[\int_0^\infty \left[\frac{\varphi_{i\omega}}{p_\omega(a)}\right]^{\sigma-1} N_r \cdot h(a)da\right]^{\frac{1}{1-\sigma}} = N_r^{\frac{1}{1-\sigma}} \frac{p(\tilde{a})}{\varphi_j(\tilde{a})} \tag{4.12}$$

where $\tilde{a} = \left[\int_0^\infty \frac{\phi(a)^{(\sigma-1)(\gamma_i-\xi)}}{\phi(\tilde{a})^{(\sigma-1)(\gamma_i-\xi)}} a^{\sigma-1} h(a)da\right]^{\frac{1}{\sigma-1}}$ is the weighted average of firm productivity level. Then the total revenue of all grocery firms in county r is:

$$R_r = \int_{\omega\in\Omega_{ri}} R_{r\omega} d\omega = \int_0^\infty R_{r\omega}(a) \cdot N_r \cdot h(a)da = N_r \cdot R_{r\omega}(\tilde{a}) \tag{4.13}$$

Recall that we have discussed earlier that under monopolistic competition, firms will enter the market until the last firm gets zero profit. Then the cutoff productivity level of firms who would enter the market is $a^* = \inf\{a : \Pi(a) > 0\}$. This means that any firm drawing a productivity level $a < a^*$ will immediately exit and never produce. Thus, $h(a)$ is a conditional distribution of $g(a)$ on $[a^*, \infty)$:

$$h(a) = \begin{cases} \frac{g(a)}{1-G(a^*)} \text{ if } a \geq a^* \\ 0 \text{ otherwise} \end{cases} \tag{4.14}$$

where $1 - G(a^*)$ is the ex-ante probability of successful entry. Then the weighted average productivity level $\tilde{a}$ can be defined as a function of cutoff level a^*:

$$\tilde{a} = \left[\frac{1}{1-G(a^*)} \times \int_{a^*}^\infty \frac{\phi(a)^{(\sigma-1)(\gamma_j-\xi)}}{\phi(\tilde{a})^{(\sigma-1)(\gamma_j-\xi)}} a^{\sigma-1} g(a)da\right]^{\frac{1}{\sigma-1}} \tag{4.15}$$

Under the zero profit condition, the number of firms in a county r in the equilibrium is determined as follows.

$$N_r = \frac{R_r}{\bar{R}_{r\omega}} = \frac{R_r}{R_{r\omega}(\tilde{a})} \tag{4.16}$$

where $\bar{R}_{r\omega}$ is the average revenue of a firm.

4.3 Model and Empirical Facts

In this sub-section, I show how the model and conditions on the parameters could rationalize the two main empirical findings via the two channels.

Fact 1: High-inequality counties have more varieties. We cannot find a tractable expression for N_r from the equation (4.16), but we could find the relationship between income inequality and the number of firms in a region as $\frac{\partial \log(N_r)}{\partial \text{Gini}_r} = \frac{\partial \log(N_r)}{\partial \delta} \cdot \frac{\partial \delta}{\partial \text{Gini}_r}$. In order to capture Fact 1, we need conditions on the parameters so that $\frac{\partial \log(N_r)}{\partial \text{Gini}_r} > 0$. Under the assumption of log-normal income distribution, we have $\frac{\partial \delta}{\partial \text{Gini}_r} > 0$. Hence, finding conditions on the parameters to have $\frac{\partial \log(N_r)}{\partial \text{Gini}_r} > 0$ is equivalent to finding conditions on the parameters to have $\frac{\partial \log(N_r)}{\partial \delta}$.

Proposition 1. *If and only if* $0 < \xi < \min\{\gamma_P, \gamma_R\}$, $\gamma_P < \gamma_R$, *and* $\beta > 0, b > 0$, *then we have:*

$$\frac{\partial \log(N_r)}{\partial Gini_r} > 0$$

Proof. See Appendix C. ■

The first condition $0 < \xi < \min\{\gamma_P, \gamma_R\}$ means that an increase in marginal costs needs not to exceed consumer valuation for quality. The second condition $\gamma_P < \gamma_R$ implies that the rich have higher quality valuation than the poor. Finally, the third condition that $\beta > 0, b > 0$ implies higher quality entails higher fixed cost and larger sales induce higher optimal quality.

The intuition for Fact 1 starts from the point that there is difference in quality valuation between different income groups. We have the rich have higher quality valuation than the poor. In high-inequality counties, the range of household income is wider. Thus, if a firm expects to sell more to the rich, then they will produce a higher quality product. Similarly, if a firm expects to sell more to the poor, then they will produce a lower quality product. It opens up the quality space of products in high-inequality areas, and thus, it allows more firms with lower productivity, who produce lower and middle quality products could enter the market. Whereas in low-inequality counties there is an intense competition among firms as households have similar income level, which makes the market tight and it is harder for entrants to enter the market.

Fact 2: Households in high-inequality county consume fewer varieties. This fact requires the conditions on parameters so that $\frac{\partial \log(|\Omega_{ri}|)}{\partial Gini_r} < 0$. We notice that $\frac{\partial \log(|\Omega_{ri}|)}{\partial Gini_r} = \frac{\partial \log(|\Omega_{ri}|)}{\partial \log(N_r)} \cdot \frac{\partial \log(N_r)}{\partial Gini_r}$

where $\frac{\partial \log(N_r)}{\partial Gini_r} > 0$ with the conditions on the parameters as in Fact 1. Therefore, to capture Fact 2, the conditions on the parameters to have $\frac{\partial \log(|\Omega_{ri}|)}{\partial Gini_r} < 0$ are equivalent to the conditions on the parameters to have $\frac{\partial \log(|\Omega_{ri}|)}{\partial \log(N_r)} < 0$.

Proposition 2. *If and only if $\sigma > 1$, $\theta > \sigma - 1$, $\varepsilon > \frac{1}{\sigma-1}$, $\varepsilon_j(j) < 0$, $\varepsilon\_j(j) > 0$, and $|\varepsilon_j(j)| < |\varepsilon_{-j}(j)|$, then we have:*

$$\frac{\partial \log(|\Omega_{ri}|)}{\partial \log(N_r)} = \frac{1}{z}\left[\frac{1}{\theta} - \varepsilon_j(j)\frac{N_r}{N_{rj}} - \varepsilon_{\_j}(j)\frac{N_r}{N_{r\_j}}\right] < 0$$

where $z = \varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$

Proof. See Appendix C. ■

The first condition $\sigma > 1$ means households find it hard to substitute one variety for the others. It supports the ideal-variety channel that households prefer to consume their favourite products. A larger θ in the second condition with $\theta > \sigma - 1$ makes the distribution of price-adjusted taste flatter. $\varepsilon > \frac{1}{\sigma-1}$ means that the cost of consuming varieties increases with the number of consumed varieties. It ensures that households would not consume all varieties available, which matches the data. Together, all these conditions make $z > 0$ and ensure the existence of interior solution for the optimal number of varieties that household consumes. Moreover, we also need the elasticity of the cost of consuming varieties with respect to the number of available varieties targeted for households type j, $\varepsilon_j(j)$, to be negative; while the elasticity of the cost of consuming varieties with respect to the number of available varieties less targeted for households type j, $\varepsilon\_j(j) > 0$, to be positive. Finally, we need $|\varepsilon_j(j)| < |\varepsilon_{-j}(j)|$, which means households are overwhelmed by bad varieties (that is, the varieties less targeted for them) more than they appreciate better products (that is, the varieties targeted for them). In other words, if the number of both types of varieties doubles, household welfare decreases as they are overwhelmed and not happy with the big number of varieties less targeted for them available in the market which overshadows the products targeted for them.

Next I will show that the model and the conditions on the parameters also satisfy the two channels which could rationalize the two empirical facts.

Ideal Variety Channel: Households in high-inequality counties consume varieties which are closer to their ideal varieties. I will assess the relation between the cutoff of the price-adjusted taste parameter of household φ^*_{ri} and the number of varieties available in a county N_r. From equation

(4.7) for the cutoff of the price-adjusted taste parameters and equation (4.9) for the optimal number of varieties household consumes, we have:

$$\frac{\partial \log\left(\tilde{\varphi}^*_{ri}\right)}{\partial \log\left(N_r\right)} = \frac{1}{\theta} - \frac{1}{\theta} \cdot \underbrace{\frac{1}{z}\left[\frac{1}{\theta} - \varepsilon_j(j)\frac{N_r}{N_{rj}} - \varepsilon_{\_j}(j)\frac{N_r}{N_{r\_j}}\right]}_{\frac{\partial \log(|\Omega_{ri}|)}{\partial \log(N_r)}<0} > 0 \quad \text{where} \quad \tilde{\varphi}^*_{ri} = \min_{\omega \in \Omega_{ri}} \{\tilde{\varphi}_{i\omega}\} \tag{4.17}$$

We can see that with conditions on parameters for Fact 2, we immediately have $\frac{\partial \gamma^*_{r,i}}{\partial N_r} > 0$, which means that as the number of varieties offered in a region increases, the cutoff of price-adjusted taste parameter will increase. This implies that households now consume products with higher price-adjusted taste parameters, so they consume products closer to their ideal varieties.

Market Segmentation Channel: Households in high-inequality counties consume more varieties if there are more varieties targeted for their type. From equation (4.9) for the optimal number of varieties a household consumes, we have:

$$\frac{\partial \log\left(|\Omega_{rj}|\right)}{\partial \log\left(N_{rj}\right)} = \frac{1}{z}\left[\frac{1}{\theta}\frac{N_{rj}}{N_r} + \underbrace{\varepsilon_{\_j}(j)}_{>0}\frac{N_{rj}}{N_{r\_j}} - \underbrace{\varepsilon_j(j)}_{<0}\right] > 0 \tag{4.18}$$

Similarly, the conditions on parameters for Fact 2 gives us $\frac{\partial \log(|\Omega_{rj}|)}{\partial \log(N_{rj})} > 0$. It implies that as the number of varieties targeted for a household type increases, the more varieties households of that type would purchase.

5 Parameter Estimation

This section presents the parameter estimation and model calibration. Since most of the parameters in my model have not been estimated in the literature yet, I use the NielsenIQ data and equations from the model to estimate them.

5.1 Elasticity of Substitution σ

I first estimate the demand elasticity σ. From equation (4.4), we get the following estimation equation:

$$\Delta \log \left(s_{ri\omega t}\right) = (1-\sigma)\Delta \log \left(p_{\omega t}\right) + \text{ Fixed effects } + \varepsilon_{ri\omega t} \tag{5.1}$$

where r, i, ω, and t denote county, household income groups, UPC, and quarters or half years. $s_{ri\omega t}$ are budget shares of variety ω. Fixed Effects are household group-by-county-by-quarter fixed effects in the regression using quarterly data, or household group-by-county-by-half-year fixed effects in the regression using half-year data.

I use the data at household group-by-county-by-half-year or household group-by-county-by-quarter in the estimation of σ. There are several reasons for this change in the sample data used to estimate σ compared to the data aggregation method used in the empirical part in section 3. First, to avoid the complication of using different products as the base product for different households, I exploit the panel dimension of the data. Specifically, I calculate the change of budget share of the same product in household consumption basket over time, instead of the ratio of budget share of two different products in household consumption basket at the same time period. Second, when taking the ratio of budget share of products over the time, another concern might arise that households' grocery expenditure and the price index are not the same for a household in two different years. To address this concern, the higher frequency data we could use, the more similar these two elements, and thus, it is more reasonable if we cancel these two terms when taking the ratio. Therefore, I use higher frequency data, that is, quarterly and half-year data instead of yearly data to ensure the accuracy of the estimates. Third, we need to calculate the ratio of budget share by product in two consecutive time period (quarter or half-year), but a household might not purchase the same product over the period, and thus, it creates many missing data for the ratio of budget share. To address this issue, instead of using individual household data, I aggregate expenditure share of households by income group since we are interested in the effects of variety inequality for different income groups.

To address the concerns about autocorrelation in the error term, I cluster standard errors at the county level. To address the simultaneity concern that taste shocks in the error term are correlated with observed price changes, I make the identifying assumption that consumer taste shocks are idiosyncratic across counties whereas supply-side cost shocks are correlated across space, following the literature in industrial organization (e.g., Hausman, 1999; Hausman and Leibtag, 2007). For the supply-side variation needed to identify σ, I follow DellaVigna and Gentzkow (2019) and Faber and Fally (2022) and exploit the fact that store chains frequently price nationally or regionally without taking into consideration changes in local demand conditions. I use either national or state-level leave-out mean price changes of UPCs as the instrument for local consumer price changes. These two instruments identify potentially different local average treatment effects. The national leave-out means instrument estimates σ for retail chains which price their products nationally, while the state leave-out means instrument could extend the complier group to regional retailers.

Table 7 shows the estimation results using OLS specification (Panel A), IV estimation using national-level instrument (Panel B), IV estimation using state-level instrument (Panel C), and IV estimation using both national and state-level instrument (Panel D) when using half-year and quarterly data. I also try different percentile dividers for income group.[12] I find that the estimates change from positive in the OLS specification to negative and statistically significant in IV estimations. The estimates from IV regressions suggest that the values of the elasticity of substitution are in the range of 1.4 to 1.75 when defining variety as UPC. Similarly, Table 16 shows the estimation results when defining variety as brand, and the values of the elasticity of substitution are in the range of 2.4 to 2.75. These estimates are very close to the result obtained in Faber and Fally (2022) who use brand-level data. They also fall at the centre of a large existing literature in IO and Marketing using brand level consumption data to estimate the sales-to-price elasticity of demand. They are, however, lower than the value of σ in Neiman and Vavra (2023), but they do not estimate this parameter. As explained above, the higher-frequency the data, the better the estimates. Thus, I use the value of the estimate from the regression using quarterly data and both national and state instruments in the analysis in the next part.

[12]To be precise, the income group here refers to the group by household grocery expenditure. To address the concern that using household grocery expenditure can be misleading as many other factors of household characteristics are correlated, I first regress household grocery expenditure on household characteristics, like household size, household composition, gender, education, occupation of the head of household. Then I use the residual of this regression to split the data by percentiles.

5.2 Price-adjusted taste Pareto distribution

In the next step, I estimate the parameters of the price-adjusted taste Pareto distribution. They are the parameter θ which governs the curvature of the distribution and the parameter φ_m which is the minimum possible value of price-adjusted taste.

Curvature of the distribution θ. From equation (4.7) of the cutoff threshold of the price-adjusted taste parameters, we take log both sides and obtain the following estimation equation.

$$\log\left(\frac{|\Omega_{ri}|}{N_r}\right) = \underbrace{\theta \log\left(\varphi_m\right)}_{\text{constant term}} - \theta \log\left(\tilde{\varphi}^*_{ri}\right) + \text{Fixed effect} + \varepsilon_{ri} \tag{5.2}$$

First, from the result of the regression (5.1), we obtain the residual, which is the taste of household group for products $\varphi_{i\omega}$. Then we can calculate $\varphi^*_{ri} = \min_{\omega\in\Omega_{ri}}\left\{\tilde{\varphi}_{i\omega}\right\}$ which is the minimum value of price-adjusted tastes of the last product that a household group would consume. Fixed effects are household group-by-county. We obtain the value for θ to be 1.802 as reported in Table 8.

Minimum value of the price-adjusted taste in the distribution φ_m. The estimate of this parameter can be obtained by backing out its value from the constant term in the estimation 5.2. Again, the value is reported in Table 8.

5.3 Estimation of z

The next parameter we will estimate is a combination of three parameters $z = \varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$. The equation we use to estimate z is derived from the solution of household's first-stage problem. The detailed derivation is presented in the appendix C

$$\log\left(\frac{\text{Var}\left(|\Omega_{ri}|\right)}{\left(|\Omega_{ri}|\right)^2}\right) = \log\left(\text{Var}\left(\left(E_{ri} - C_{ri0}\right)^{\frac{1}{z}}\right)\right) - \frac{2}{z}\log\left(E_{ri} - C_{ri0}\right) \tag{5.3}$$

where $z = \varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$. The left-hand side of this equation and household grocery expenditure $E_{ri} - C_{ri0}$ are obtained directly from the data. To ensure good variation in the data when taking the variance, I return to using yearly household level data, and take the variance across households within a county each year. Then I use GMM to find the value of z that minimize this moment. Using barcode level data, I find the value of z to be 1.098, as reported in Table 8.

5.4 Cost of consuming varieties

In the next step, I estimate the set of parameters associated with the cost of consuming varieties. They are the elasticity of the cost of consuming varieties with respect to the number of varieties household consume (ε), the elasticity of the cost of consuming varieties with respect to the number of varieties targeted for household $\varepsilon_j(j)$, the elasticity of the cost of consuming varieties with respect to the number of varieties less targeted for household $\varepsilon_{\_j}(j)$, and the fixed cost (F).

Elasticity ε. Recall that in the previous part we have obtained the estimates of $z = \varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$. We also have the estimates of θ and σ. Thus, we can back out the value of ε. The result of the estimate is reported in Table 8.

Elasticity $\varepsilon_j(j), \varepsilon_{\_j}(j)$. From equation (4.9) for the optimal number of varieties household consume, we can take log both sides and obtain the following equation.

$$\log\left(|\Omega_{ri}|\right) = \underbrace{A}_{\text{constant term}} + \frac{1}{\theta \cdot z}\log\left(N_{rj} + N_{r\_j}\right) - \frac{\varepsilon_j(j)}{z}\log\left(N_{rj}\right) - \frac{\varepsilon_{\_j}(j)}{z}\log\left(N_{r\_j}\right) \tag{5.4}$$

where $A = \frac{(E_{ri}-C_{ri0})\varphi_m}{\varepsilon F}\left(\frac{1}{\sigma-1} - \frac{1}{\theta}\right)\left(1 + \frac{1-\sigma}{\theta}\right)^{\frac{1}{1-\sigma}}$. Then we can use GMM method to obtain the estimated values for $\varepsilon_j(j)$ and $\varepsilon_{\_j}(j)$ for two types of households: the rich and the poor. To identify the number of varieties targeted for the rich and the number of varieties targeted for the poor, I make the assumption that high-quality products are those targeted for the rich, and lower-quality products are those targeted for the poor. This assumption relies on the empirical findings from Handbury (2021) that rich households tend to consume higher quality products. Within each product module, I define high-quality products as those whose prices are above the mean or the median price of all prices in the module, and the rest is low-quality products. I define them within a product module to ensure the accuracy in comparing products as varieties in different module might have very different price levels. The values of the estimates are reported in Table 8 for both the rich and the poor.

Fixed cost F. Finally, we obtain the estimated value for the fixed cost of consuming varieties F by using the equation (4.9) for the solution of optimal number of varieties. The value is reported in Table 8.

5.5 Model Replicates Empirical Facts

I do not use the results from section 3 of the empirical parts in estimating model parameters. Thus, we could use the estimated parameters and equations from the model to calculate the equivalence of the empirical moments and compare them to check how well our model could replicate the empirical facts. Table 9 shows that the model can replicate the empirical facts well. The 95% confidence interval from the model is calculated using the 95% confidence interval of all estimated parameters that we have their confidence interval.

6 Welfare Implications

We are interested in the difference in household welfare between when they live in a low-inequality county and when they live in a high-inequality county. In this section, I use the model and the estimated parameters to explore the welfare implications of variety inequality. To assess welfare changes, we could measure either the change in household's indirect utility or the change in household's consumption (i.e., consumption equivalence).

Indirect Utility function. The most direct method of assessing welfare impacts is to compare the changes in household's indirect utility when the household is in low-inequality county and then move to a higher-inequality county. From the equation (4.5) of household's indirect utility and substituting the equation (4.8) for the price index, we obtain the following equations for the indirect utility of the poor:

$$\boldsymbol{U_{rP}} = \frac{\left(E_{ri} - C_{ri0}\right)\varphi_m}{\left(|\Omega_{ri}|\right)^{\frac{1}{1-\sigma}+\frac{1}{\theta}}\left(1+\frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}}\left(\frac{1}{N_r}\right)^{\frac{1}{\theta}}} - F \times \left(|\Omega_{\mathrm{ri}}|\right)^{\varepsilon} N_{rP}^{\varepsilon_P(P)} N_{rR}^{\varepsilon_R(P)} \tag{6.1}$$

and the rich:

$$\boldsymbol{U_{rR}} = \frac{\left(E_{ri} - C_{ri0}\right)\varphi_m}{\left(|\Omega_{ri}|\right)^{\frac{1}{1-\sigma}+\frac{1}{\theta}}\left(1+\frac{1-\sigma}{\theta}\right)^{\frac{1}{\sigma-1}}\left(\frac{1}{N_r}\right)^{\frac{1}{\theta}}} - F \times \left(|\Omega_{\mathrm{ri}}|\right)^{\varepsilon} N_{rR}^{\varepsilon_R(R)} N_{rP}^{\varepsilon_P(R)} \tag{6.2}$$

I consider both types of households when living in low-inequality, median-inequality, and high-inequality counties. Since I am only interested in the dimension of income inequality, before classifying counties by their income inequality level into different bins, I regress county-level Gini index on the set of county-level controls used in the empirical section and use the residual from this regression to group counties by quantiles. Then we can calculate the average number of varieties, the average number of varieties targeted for households of each type, the average number of varieties less targeted for households of each type by income inequality quantiles. Finally, we can calculate the changes in household welfare in counties with different level of income inequality.

Table 10 - Panel A shows the results using my estimated parameters from section 5. Both the poor and the rich experience a welfare loss when living in higher-inequality counties. It implies that the market segmentation channel dominates the ideal variety channel. In addition, quantitatively we observe that the rich are negatively affected more when living in higher-inequality counties.

Instead of using the estimated parameters, we could use some parameters from the literature to evaluate household welfare. Since the setup of my model is closely related to the model setup in Neiman and Vavra (2023), I replace my estimated parameters with parameters from Neiman and Vavra (2023) whenever possible. For some parameters which are specific to my models, I keep using my estimates. In particular, I use the following estimates from Neiman and Vavra (2023): $\sigma = 4.7$, $\theta = 7.9$, $\varepsilon = 2$, $\varphi_m = 1$, and $F = 0.055$. Table 10 - Panel B presents this result. We can see that using the literature estimates gives us strange results: the effects are in different directions. This implies that using literature estimates would not give us the results that are consistent with the data, and thus, estimating these parameters using data and model equations is the suitable method to assess welfare implications of variety inequality.

Consumption Equivalence. Another approach to compare household welfare when living in different income-inequality counties is consumption equivalence. That is we are interested in how much consumption a household loses or gains when living in higher-inequality county. Table 10 - Panel C shows the results of this approach using my estimates. The result is consistent with the indirect utility approach: the market segmentation channel dominates the ideal variety channel. Hence, households are worse off when living in higher-inequality counties. We also observe a similar pattern that the rich's consumption decreases more than the poor when living in higher-inequality counties

7 Conclusion

This paper studies how local income inequality leads to different household consumption patterns through the differences in the scope of products offered by retailers in different regions. To do so, I exploit NielsenIQ's detailed home and store scanner datasets which have well-defined concepts of product variety. These datasets enable us to trace the set of products available in a county and the set of products purchased by households. I leverage these data to explore the underlying channels and propose a model with an endogenous choice of variety by both households and firms to rationalize the observed empirical facts and quantify the impacts on household welfare.

I document two novel findings showing the relationship between income inequality and variety inequality. First, high-inequality counties have more varieties. Paradoxically, households within these high-inequality counties consume fewer varieties. The same patterns hold when we focus on people who move from one place to another.

Subsequent examination of extensive and intensive margin effects reveals two plausible channels that may account for these intriguing empirical findings. The first channel, referred to as the "ideal-variety channel," posits that high-inequality regions offer a greater array of product varieties, which makes it easier for households to find products closer to their ideal varieties. Consequently, they tend to concentrate their expenditures on these favored selections. The second channel, the "market segmentation channel," posits that in high-inequality counties, more firms are inclined to produce goods catering to the two ends of the income distribution. Thus, despite an overall increase in the total number of product varieties, fewer options are available to each group of households. If households prefer products aligned with their quality expectations, the reduced availability of such options within their quality segment compels them to consume a more limited range of varieties.

Both channels coexist in the data, yet they yield different implications for household welfare. To quantify the welfare impact of income inequality through the variety channel and assess which channel predominates, I develop a model with an endogenous choice of variety by both sides of the market. The model explains the puzzling facts through two features of household utility and firm technology. On the demand side, households find it harder to substitute products. Moreover, richer households value higher quality attributes more compared to poorer households. On the supply side, an increase in a firm's marginal cost needs not surpass consumer valuation for quality. Producing higher product quality requires a higher fixed cost of production.

Using the estimated parameters from the model, I find that the "market segmentation channel" prevails over the "ideal-variety channel". Consequently, households experience a decline in their welfare when residing in higher-inequality areas. My findings suggest that having access to more varieties does not necessarily mean household welfare would improve as the predictions in the love-of-variety model. My results also supports the recent literature on household consumption concentration.Neiman and Vavra (2023) use aggregation data to show that households have been increasingly concentrating their consumption on niche products, I explore cross-sectional variation across regions to show the impact of income inequality on household consumption.

References

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

— **and Mark Bils**, "Has Consumption Inequality Mirrored Income Inequality?," *American Economic Review*, 2015, *105* (9), 2725–56.

Algert, Susan J., Aditya Agrawal, and Douglas S. Lewis, "Disparities in Access to Fresh Produce in Low-Income Neighborhoods in Los Angeles," *American Journal of Preventive Medicine*, November 2006, *30* (5), 5–370.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell, "Food Deserts and the Causes of Nutritional Inequality," *The Quarterly Journal of Economics*, 2019, *134* (4), 1793–1844.

Alwitt, Linda F. and Thomas D. Donley, "Retail Stores in Poor Urban Neighborhoods," *The Journal of Consumer Affairs*, November 1997, *31* (1), 139–164.

Amiti, Mary and Amit K. Khandelwal, "Import Competition and Quality Upgrading," *The Review of Economics and Statistics*, May 2013, *95* (2), 476–490.

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

Attanasio, Orazio and Luigi Pistaferri, "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure," *American Economic Review*, 2014, *104* (5), 122–26.

Attanasio, Orazio P. and Luigi Pistaferri, "Consumption Inequality," *Journal of Economic Perspectives*, May 2016, *30* (2), 3–28.

Baker, Elizabeth, Mario Schootman, Ellen Barnidge, and Cheryl Kelly, "The Role of Race and Poverty in Access to Foods That Enable Individuals to Adhere to Dietary Guidelines," *Preventing chronic disease*, 08 2006, *3*, A76.

Benassi, Corrado, Alessandra Chirco, and Caterina Colombo, "Vertical Differentiation And The Distribution Of Income," *Bulletin of Economic Research*, October 2006, *58* (4), 345–367.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott, "Firms in International Trade," *Journal of Economic Perspectives*, September 2007, *21* (3), 105–130.

Berry, Steven, "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25* (2), 242–262.

Bertrand, Marianne and Adair Morse, "Trickle-Down Consumption," *The Review of Economics and Statistics*, December 2016, *98* (5), 863–879.

Besharov, Douglas J., Marianne Bitler, and Steven J. Haider, "An economic view of food deserts in the united states," *Journal of Policy Analysis and Management*, December 2011, *30* (1), 153–176.

Chancel, Lucas and Thomas Piketty, "Global Income Inequality, 1820–2020: the Persistence and Mutation of Extreme Inequality [Global Inequality of Hourly Income, 1980–2020]," *Journal of the European Economic Association*, 2021, *19* (6), 3025–3062.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, " Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *," *The Quarterly Journal of Economics*, 09 2014, *129* (4), 1553–1623.

Christen, Markus and Ruskin Morgan, "Keeping Up With the Joneses: Analyzing the Effect of Income Inequality on Consumer Borrowing," *Quantitative Marketing and Economics (QME)*, June 2005, *3* (2), 145–173.

Coibion, Olivier, Yuriy Gorodnichenko, and Dmitri Koustas, "Consumption Inequality and the Frequency of Purchases," *American Economic Journal: Macroeconomics*, 2021, *13* (4), 449–82.

— **,** — **, and Gee Hee Hong**, "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications," *American Economic Review*, March 2015, *105* (3), 993–1029.

— , — , **Marianna Kudlyak, and John Mondragon**, "Does Greater Inequality Lead to More Household Borrowing? New Evidence from Household Data," *Journal of European Economics Association*, 2020, *18*, 2922–2971.

Cravino, Javier and Andrei A. Levchenko, "The Distributional Consequences of Large Devaluations," *American Economic Review*, November 2017, *107* (11), 3477–3509.

— **, Ting Lan, and Andrei A. Levchenko**, "Price stickiness along the income distribution and the effects of monetary policy," CEPR Discussion Papers 12967, C.E.P.R. Discussion Papers June 2018.

Daly, Mary and Daniel Wilson, "Keeping up with the Joneses and staying ahead of the Smiths: evidence from suicide data," Working Paper Series 2006-12, Federal Reserve Bank of San Francisco 2006.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform Pricing in U.S. Retail Chains," *The Quarterly Journal of Economics*, 06 2019, *134* (4), 2011–2084.

Desmet, Klaus and Romain Wacziarg, "Understanding Spatial Variation in COVID-19 across the United States," *Journal of Urban Economics, vol. 127, January 2022, 103332.*, January 2022, *127* (103332).

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, June 1977, *67* (3), 297–308.

Dorfman, Robert, "A Formula for the Gini Coefficient," *The Review of Economics and Statistics*, 1979, *61* (1), 146–49.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, September 2011, *79* (5), 1453–1498.

Faber, Benjamin and Thibault Fally, "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," *Review of Economic Studies*, 2022, *89* (3), 1420–1459.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Sources of Geographic Variation in Health Care: Evidence From PatientMigration," *The Quarterly Journal of Economics*, 2016, *131* (4), 1681–1726.

George, Lisa and Joel Waldfogel, "Who Affects Whom in Daily Newspaper Markets?," *Journal of Political Economy*, 2003, *111* (4), 765–784.

Gilbert, Laurien, "Gains from Product Variety and the Local Business Cycle," *Job Market Paper*, November 2017.

Hallak, Juan and Jagadeesh Sivadasan, "Product and process productivity: Implications for quality choice and conditional exporter premia," *Journal of International Economics*, 2013, *91* (1), 53–67.

Handbury, Jessie, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," *Econometrica*, 2021, *89*, 2679–2715.

— **and David E. Weinstein**, "Goods Prices and Availability in Cities," *Review of Economic Studies*, 2015, *82* (1), 258–296.

Hausman, Jerry, "Cellular Telephone, New Products, and the CPI," *Journal of Business Economic Statistics*, 1999, *17* (2), 188–94.

— **and Ephraim Leibtag**, "Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart," *Journal of Applied Econometrics*, 2007, *22* (7), 1157–1177.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni Violante, "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006," *Review of Economic Dynamics*, 2010, *13* (1), 15–51.

Heffetz, Ori, "A Test of Conspicuous Consumption: Visibility and Income Elasticities," *The Review of Economics and Statistics*, November 2011, *93* (4), 1101–1117.

Helpman, Elhanan and Paul Krugman, *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Vol. 1 of *MIT Press Books*, The MIT Press, August 1985.

Horowitz, C.R., K.A. Colson, P.L. Hebert, and K. Lancaster, "Barriers to buying healthy foods for people with diabetes: Evidence of environmental disparities," *American Journal of Public Health*, 2004, *94* (9), 1549–1554.

Hsu, Wen-Tai, Lin Lu, and Pierre Picard, "Income Inequality, Productivity, and International Trade," Economics and Statistics Working Papers 13-2018, Singapore Management University, School of Economics July 2018.

Hummels, David and Peter J. Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review*, June 2005, *95* (3), 704–723.

Hyun, Jungsik and Ryan Kim, "Spillovers and Redistribution through Intra-Firm Networks: The Product Replacement Channel," *Working paper*, November 2019.

Jappelli, Tullio and Luigi Pistaferri, "Does Consumption Inequality Track Income Inequality in Italy?," *Review of Economic Dynamics*, January 2010, *13* (1), 133–153.

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *The Quarterly Journal of Economics*, December 2018, *134* (2), 715–783.

Jaskold, J. Gabszewicz and J. F. Thisse, "Price competition, quality and income disparities," *Journal of Economic Theory*, June 1979, *20* (3), 340–359.

— **and** — , "Entry (and exit) in a differentiated industry," *Journal of Economic Theory*, April 1980, *22* (2), 327–338.

Jetter, Karen and Diana Cassady, "The Availability and Cost of Healthier Food Alternatives," *American journal of preventive medicine*, 02 2006, *30*, 38–44.

Kaplan, Greg and Sam Schulhofer-Wohl, "Inflation at the household level," *Journal of Monetary Economics*, 2017, *91* (C), 19–38.

Karpyn, Allison, Danielle Riser, Tara Tracy, Rui Wang, and Yi Shen, "The changing landscape of food deserts.," *UNSCN nutrition*, 2019, *44*, 46–53.

Khandelwal, Amit, "The Long and Short (of) Quality Ladders," *The Review of Economic Studies*, 10 2010, *77* (4), 1450–1476.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Krueger, Dirk and Fabrizio Perri, "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory -super-1," *Review of Economic Studies*, 2006, *73* (1), 163–193.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, December 1980, *70* (5), 950–959.

Krugman, Paul R., "Increasing returns, monopolistic competition, and international trade," *Journal of International Economics*, November 1979, *9* (4), 469–479.

Kugler, Maurice and Eric Verhoogen, "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, 2012, *79* (1), 307–339.

Lancaster, Kelvin, *Variety, Equity, and Efficiency: Product Variety in an Industrial Society*, New York Chichester, West Sussex: Columbia University Press, 1979.

— , "Competition and Product Variety," *The Journal of Business*, 1980, *53* (3), S79–103.

Larson, Nicole, Mary Story, and Melissa Nelson, "Neighborhood Environments:: Disparities in Access to Healthy Foods in the US," *American journal of preventive medicine*, 11 2008, *36*, 74–81.

Lee, Chulhee, "Rising family income inequality in the United States, 1968-2000: impacts of changing labor supply, wages, and family structure," *International Economic Journal*, 2008, *22* (2), 253–272.

Li, Nicholas, "An Engel Curve for Variety," *The Review of Economics and Statistics*, March 2021, *103*

(1), 72–87.

Ludwig, Jens, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg Duncan, Lawrence Katz, Ronald Kessler, Jeffrey Kling, Stacy Lindau, Robert Whitaker, and Thomas Mcdade, "Neighborhoods, Obesity, and Diabetes — A Randomized Social Experiment," *The New England journal of medicine*, 10 2011, *365*, 1509–19.

Luttmer, Erzo F. P., "Neighbors as Negatives: Relative Earnings and Well-Being," Working Paper Series rwp04-029, Harvard University, John F. Kennedy School of Government July 2004.

Malmendier, Ulrike and Leslie Sheng Shen, "Scarred Consumption," NBER Working Papers 24696, National Bureau of Economic Research June 2018.

Maurer, Jürgen and André Meier, "Smooth it Like the 'Joneses'? Estimating Peer-Group Effects in Intertemporal Consumption Choice," *Economic Journal*, March 2008, *118* (527), 454–476.

Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Meyer, Bruce D. and James X. Sullivan, "Consumption and Income Inequality in the United States since the 1960s," *Journal of Political Economy*, 2023, *131* (2), 247 – 284.

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

Neumark, David and Andrew Postlewaite, "Relative income concerns and the rise in married women's employment," *Journal of Public Economics*, October 1998, *70* (1), 157–183.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, 2003, *118* (1), 1–41.

Powell, Lisa, Sandy Slater, Donka Mirtcheva, Yanjun Bao, and Frank Chaloupka, "Food Store Availability and Neighborhood Characteristics in the United States," *Preventive medicine*, 04 2007, *44*, 189–95.

Reed, Deborah and Maria Cancian, "Rising Family Income Inequality: The Importance of Sorting," Mathematica Policy Research Reports 2012.

Shaked, Avner and John Sutton, "Relaxing Price Competition Through Product Differentiation," *Review of Economic Studies*, 1982, *49* (1), 3–13.

Sharkey, Joseph, Scott Horel, and Wesley Dean, "Neighborhood deprivation, vehicle ownership, and potential spatial access to a variety of fruits and vegetables in a large rural area in Texas," *International journal of health geographics*, 05 2010, *9*, 26.

Smets, Vincent, Jeroen Cant, and Stefanie Vandevijvere, "The Changing Landscape of Food Deserts and Swamps over More than a Decade in Flanders, Belgium," *International Journal of Environmental Research and Public Health*, 2022, *19* (21).

Thewissen, Stefan, Lane Kenworthy, Brian Nolan, Max Roser, and Tim Smeeding, "Rising Income Inequality and Living Standards in OECD Countries: How Does the Middle Fare?," *Journal of Income Distribution*, July 2018, *26* (2), 1–23.

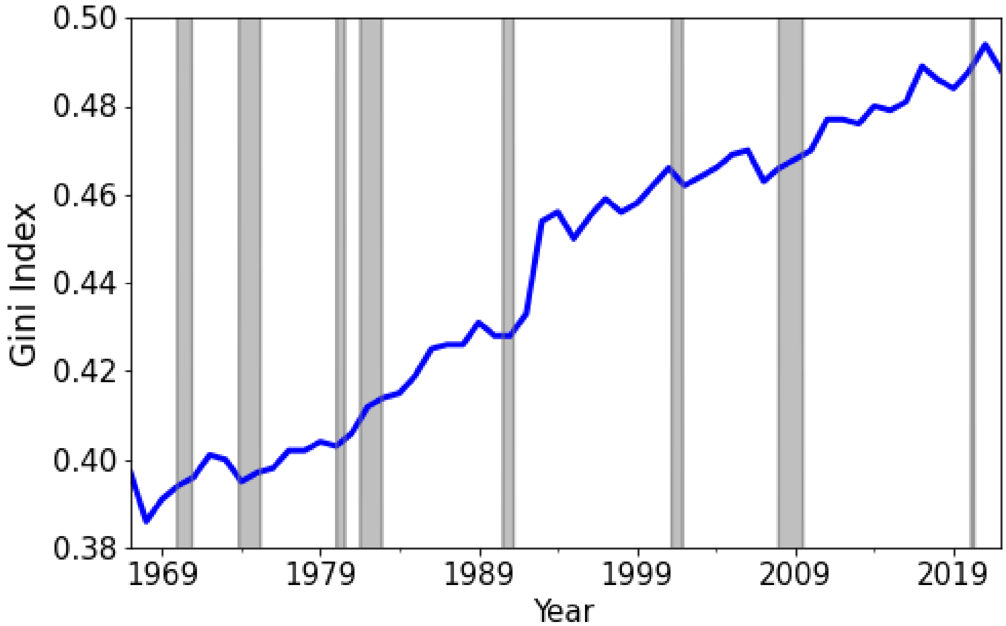
Waldfogel, Joel, "Who Benefits Whom in Local Television Markets?," *Brookings-Wharton Papers on Urban Affairs*, 2003.

— , *The Tyranny of the Market*, Harvard University Press, 2007.

Wright, James, Amy Donley, Marie Gualtieri, and Sara Strickhouser, "Food Deserts: What is the Problem? What is the Solution?," *Society*, 04 2016, *53.*

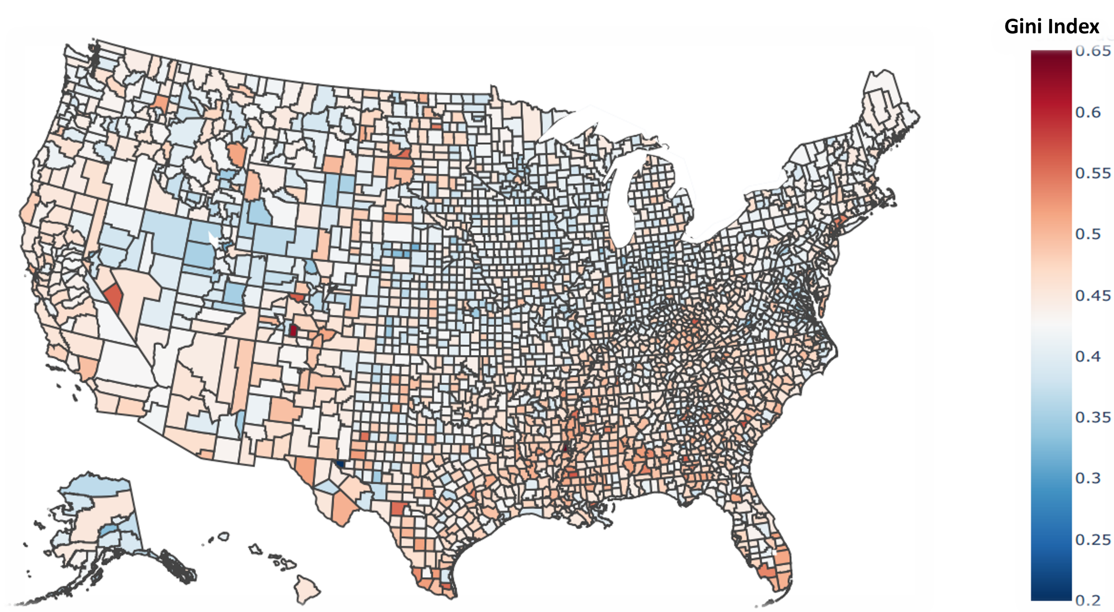
Yurko, Anna V., "How does income inequality affect market outcomes in vertically differentiated markets?," *International Journal of Industrial Organization*, July 2011, *29* (4), 493–503.

Figure 1: Income Inequality in the U.S. from 1967 to 2020



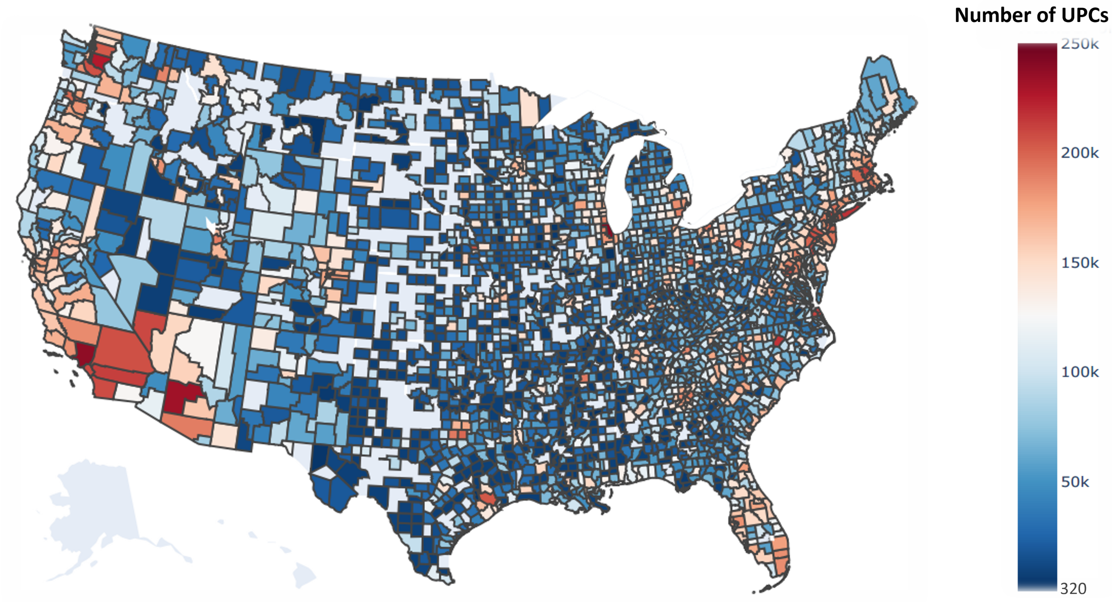
Note: The figure plots Gini index in the US from 1967 to 2020. Data is from US Census Bureau.

Figure 2: Income Inequality across U.S. counties



Note: The figure plots Gini index in 2010 at the county level. Data from American Community Survey five-years estimates 2005-2010.

Figure 3: The number of UPCs available across U.S. counties



Note: The figure plots the number of UPCs available at the county level in the U.S. in 2010. Data from NielsenIQ Retail Scanner Data.

Figure 4: Product hierarchy in NielsenIQ data

Department N = 11	⊇	Product group N ≈ 120	⊇	Product module N ≈ 1306	⊇	Brand N ≈ 252,336	⊇	UPC N ≈ 4,735,446
e.g. Frozen foods		e.g. Ice cream, novelties		e.g. Ice cream - bulk		e.g. Haagen-Dazs		e.g. Haagen-Dazs Strawberries & Cream Ice Cream 460ml
						e.g. Stonyfield		
				e.g. Ice cream – cones & cups				

Note: The figure show the product hierarchy in NielsenIQ data. From left to right, department is the most aggregate level, and UPC is the smallest level.

Table 1: Summary Statistics - Retail Scanner Data

	Number of varieties in county					**Number of stores in county**				
	Mean	St. Dev.	N	Max	Min	Mean	St. Dev.	N	Max	Min
Panel A - By Gini percentile										
$\leq 25th$ percentile	57495.41	47370.36	636	212516	316	8.121	13.814	588	126	1
$25-50th$ percentile	67162.94	49120.16	628	216782	1479	10.743	18.208	651	197	1
$50-75th$ percentile	70417.19	52284.24	632	230966	1925	16.149	36.243	678	493	1
$> 75th$ percentile	73302.46	58278.22	630	238662	2716	25.126	60.432	629	705	1
Panel B - By Population percentile										
$\leq 25th$ percentile	20562.04	14463.96	632	85053	316	1.622	.835	357	6	1
$25-50th$ percentile	38479.31	22299.29	631	115638	1479	2.503	1.461	646	9	1
$50-75th$ percentile	70342.5	28945.41	632	159819	3337	5.513	3.474	772	25	1
$> 75th$ percentile	138981.8	34782.69	631	238662	28414	41.595	60.568	771	705	3
Panel C - By Median household income percentile										
$\leq 25th$ percentile	38343.04	30300.48	623	203291	1925	4.905	10.150	630	191	1
$25-50th$ percentile	59055.03	43199.86	628	194393	316	10.035	20.733	632	234	1
$50-75th$ percentile	69125.63	51250.28	620	208244	1479	15.460	32.924	626	316	1
$> 75th$ percentile	100148.8	58628.27	655	238662	2788	29.502	60.231	658	705	1
Panel D - By Unemployment rate percentile										
$\leq 25th$ percentile	63454.67	53076.97	660	212516	2661	10.179	19.600	545	179	1
$25-50th$ percentile	75947.65	55859.62	623	230966	316	19.793	44.113	673	543	1
$50-75th$ percentile	71670.55	51319.56	612	238662	1479	17.746	42.144	710	681	1
$> 75th$ percentile	57641.49	46385.61	631	236534	3123	11.412	35.961	618	705	1
Panel E - By Poverty rate percentile										
$\leq 25th$ percentile	81229.2	58516.88	639	224534	1479	18.651	32.437	630	292	1
$25-50th$ percentile	73483.94	52862.84	631	230966	2723	16.883	37.408	649	493	1
$50-75th$ percentile	67314.67	50475.61	629	238662	316	16.865	51.250	679	705	1
$> 75th$ percentile	45956.74	38127.74	627	203291	1925	7.4218	19.316	588	228	1
Total	67074.28	52232.62	2526	238662	316	15.131	37.689	2546	705	1

Note: The table presents summary statistics of the number of varieties in county and the number of stores in county in 2010 at different percentiles of some location-specific controls.

Table 2: Summary Statistics - Consumer Panel Data

	Poor households					Rich households				
	Mean	St. Dev.	N	Max	Min	Mean	St. Dev.	N	Max	Min
Panel A - By Gini percentile										
$\leq 25th$ percentile	503.836	255.595	2571	1985	1	644.053	292.884	5315	2363	10
$25-50th$ percentile	493.172	258.215	2561	2449	1	618.849	288.939	5287	3087	4
$50-75th$ percentile	478.046	255.452	2562	2123	1	592.243	289.068	5311	2468	5
$> 75th$ percentile	447.054	236.508	2555	1898	6	555.303	283.997	5291	2231	5
Panel B - By Population percentile										
$\leq 25th$ percentile	499.118	257.891	2563	2123	1	652.583	289.812	5310	2296	5
$25-50th$ percentile	502.633	259.997	2565	1906	1	623.657	292.933	5355	3087	8
$50-75th$ percentile	481.018	251.339	2563	2449	2	584.878	286.073	5282	2468	4
$> 75th$ percentile	439.409	235.011	2558	1744	1	548.657	282.795	5257	2231	5
Panel C - By Median household income percentile										
$\leq 25th$ percentile	492.032	262.761	2577	2123	1	619.628	284.868	5335	2071	5
$25-50th$ percentile	491.594	249.974	2549	1792	1	603.977	292.567	5352	3087	11
$50-75th$ percentile	475.747	252.473	2561	2449	2	602.501	290.671	5230	2468	4
$> 75th$ percentile	462.891	243.211	2562	1744	1	584.307	293.211	5287	2263	5
Panel D - By Unemployment rate percentile										
$\leq 25th$ percentile	470.887	246.843	2709	1985	8	622.212	290.519	5358	3087	8
$25-50th$ percentile	483.095	254.977	2591	2449	1	602.04	290.537	5726	2363	4
$50-75th$ percentile	480.684	250.138	2465	1744	1	606.769	294.391	5120	2468	5
$> 75th$ percentile	488.380	257.994	2484	2123	1	578.152	285.079	5000	2048	5
Panel E - By Poverty rate percentile										
$\leq 25th$ percentile	482.831	248.206	2589	1985	1	613.467	293.945	5413	2263	5
$25-50th$ percentile	485.905	251.371	2546	2449	1	608.404	293.517	5208	3087	4
$50-75th$ percentile	474.155	253.598	2670	1684	2	607.341	288.237	5370	2231	8
$> 75th$ percentile	479.621	256.837	2444	2123	1	580.823	285.480	5213	2055	5
Total	480.5691	252.4757	10249	2449	1	602.6464	290.5914	21204	3087	4

Note: The table presents summary statistics of the number of varieties purchased by low- and high-income households in 2010 at different percentiles of some location-specific controls.

Table 3: Income Inequality and Varieties Available in County

		Number of UPCs in a county					
Panel A - by County	Gini	4.937***					
		(.675)					
	N	17,893					
	R^2	0.386					
		Number of categories			Number of UPCs by category		
		Department	Product Group	Product Module	by Department	by Product Group	by Product Module
Panel B - by County by Category	Gini	0.057*	0.149***	0.828***	7.447***	5.727***	4.630***
		(.021)	(.040)	(.148)	(1.050)	(.728)	(.562)
	N	17,893	17,893	17,893	74,286	2,035,213	15,905,437
	R^2	0.659	0.659	0.648	0.627	0.627	0.621
		Number of stores			Number of UPCs by store		
Panel C - by County by Store	Gini	6.836***			0.504**		
		(1.093)			(.176)		
	N	17,893			244,987		
	R^2	0.542			0.608		
		Number of categories by store			Number of UPCs by category by store		
		Department	Product Group	Product Module	by Department	by Product Group	by Product Module
Panel D - by Store by Category	Gini	0.080**	0.139*	0.397*	1.253***	0.592**	0.426*
		(.024)	(.053)	(.163)	(.173)	(.194)	(.165)
	N	244,987	244,987	244,987	2,356,515	26,464,263	175,131,402
	R^2	0.672	0.672	0.678	0.683	0.674	0.664

Note: The table presents the estimates of the specification (3.1) for the first fact — high-inequality counties have more varieties, and this results are robust along different dimensions of extensive and intensive margins. Panel A shows the most general result with dependent variable is the number of UPCs available in a county r in year t. Panel B disentangles the results into extensive and intensive margins by category — department, product group, product module. Panel C disentangles the results into extensive and intensive margins along the dimension of store. Lastly, Panel D shows the results for the extensive and intensive margins of the increase in the number of varieties offered within a store. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 4: Income Inequality and Varieties Consumed by Households

		All HHs	Poor HHs	Middle HHs	Rich HHs	All HHs	Poor HHs	Middle HHs	Rich HHs
		Number of UPCs							
	Gini	-.572**	-0.504**	-0.532**	-0.883***				
		(.235)	(.216)	(.248)	(.314)				
Panel A - by Household	R	0.220	0.204	0.224	0.198				
	N	676,269	100,562	457,344	118,363				
		Number of categories				**Number of UPCs by category**			
	Gini	-0.212***	-0.303***	-0.208***	-0.365***	-0.314***	-0.269**	-0.415***	-0.713***
Panel B - by Household		(.059)	(.094)	(.066)	(.160)	(.084)	(.094)	(.115)	(.119)
by Category	R	0.343	0.338	0.343	0.344	0.375	0.371	0.376	0.382
(Product Group)	N	676,269	100,562	457,344	118,363	43,521,849	10,017,813	23,738,644	9,765,392
		Number of stores				**Number of UPCs by store**			
	Gini	0.415**	0.901***	0.345*	0.738***	-0.985***	-1.272***	-0.809***	-1.236***
		(.184)	(.201)	(.191)	(.167)	(.369)	(.377)	(.219)	(.295)
Panel C - by Household by Store	R	0.111	0.146	0.112	0.118	0.083	0.083	0.081	0.082
	N	676,269	100,562	457,344	118,363	4,474,886	418,859	3,589,470	466,557
		Number of categories by store				**Number of UPCs by category by store**			
	Gini	-0.535***	-0.494**	-0.592***	-0.550*	-0.396**	-0.416***	-0.203**	-0.582***
Panel D - by Household		(.156)	(.249)	(.157)	(.296)	(.152)	(.135)	(0.090)	(.141)
by Store	R	0.021	0.023	0.021	0.020	0.008	0.009	0.008	0.007
by Category	N	4,474,886	418,859	3,589,470	466,557	85,716,243	8,479,842	67,815,220	9,421,181

Note: This table presents the estimates of the specification (3.2) for Fact 2 — households in high-inequality counties purchase fewer varieties, and this results are robust along different dimensions of extensive and intensive margins. Panel A shows the most general result with dependent variable is the number of varieties consumed by household i living in county r in year t. Panel B disentangles the results into extensive and intensive margins by category — product group. Panel C disentangles the results into extensive and intensive margins along the dimension of store. Lastly, Panel D shows the results for the extensive and intensive margins of the decrease in the number of varieties purchased within a store. The standard errors are clustered by county and household. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 5: Income Inequality and Varieties Consumed by Households

	Number of UPCs		
	All HHs	Poor HHs	Rich HHs
$log(\Sigma_{r,t})$	-0.008*	-0.011*	-0.015*
	(0.004)	(0.006)	(.008)
R^2	0.224	0.226	0.210
N	424,602	65,113	70,263

Note: Standard errors are clustered by county and household.

Note: This table presents the estimates of the specification (3.3) for Fact 2 — households in high-inequality counties purchase fewer varieties. However, instead of regressing on Gini index, I replace it with the log of the number of varieties available in a county. The standard errors are clustered by county and household. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 6: Varieties targeted for households and Varieties consumed by households

	Poor HHs		Rich HHs
Σ_{rPt}	0.019*	Σ_{rRt}	0.012*
	(0.010)		(0.007)
R^2	0.226	R^2	0.208
N	100,562	N	118,363

Note: The table presents estimates of specification (3.5). The standard errors are clustered by county and household. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 7: Estimation of Elasticity of Substitution σ (UPC-level Data)

	Quarterly Data			Half-year Data		
	2-quantiles	**4-quantiles**	**5-quantiles**	**2-quantiles**	**4-quantiles**	**5-quantiles**
Panel A - OLS						
$1-\sigma$	0.832***	0.796***	0.788***	0.884***	0.851***	0.838***
	(0.00937)	(0.0127)	(0.0136)	(0.00769)	(0.0093)	(0.0131)
N	16,702,448	10,515,252	8,423,384	13,632,768	9,021,192	7,385,709
Panel B – National IV						
$1-\sigma$	-0.542***	-0.399***	-0.344***	0.0260	0.0705**	0.0651**
	(0.0382)	(0.0419)	(0.0366)	(0.0207)	(0.0288)	(0.0250)
N	16,702,448	10,515,252	8,423,384	13,632,768	9,021,192	7,385,709
First Stage F-Stat.	694.8	1293	1424	784.4	1091	1402
Panel C – State IV						
$1-\sigma$	-0.713***	-0.506***	-0.438***	-0.194***	-0.101**	-0.0956*
	(0.0521)	(0.0594)	(0.0605)	(0.0494)	(0.0504)	(0.0573)
N	12,114,137	7,114,161	5,531,274	10,629,399	6,614,866	5,283,510
First Stage F-Stat.	1459	1192	902.8	597	559.7	535.7
Panel D – National and State IVs						
$1-\sigma$	-0.749***	-0.570***	-0.508***	-0.119***	-0.0454	-0.0744*
	(0.0502)	(0.0523)	(0.0510)	(0.0335)	(0.0383)	(0.0435)
N	12,114,137	7,114,161	5,531,274	10,629,399	6,614,866	5,283,510
First Stage F-Stat.	712.1	841.3	710.9	851.9	803.6	906.1

Note: The table presents estimates of the elasticity of substitution σ from the specification (5.1). The definition of variety is barcode UPC. 2-quantiles means splitting households by 2 groups, 4-quantiles means splitting households by 4 groups, and 5-quantiles means splitting households by 5 groups. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 8: Parameter Estimation

Parameter	**Value**	**Standard deviation**	**Meaning**
z	1.098	0.314	$\varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$
σ	1.570	0.052	Elasticity of substitution
θ	1.802	0.542	Curvature of price-adjusted taste distribution
φ_m	0.818		Minimum value of price-adjusted taste
ε	2.541		Elasticity of cost of consuming varieties w.r.t N_r
$\varepsilon_R(R)$	-0.610		Elasticity of cost of consuming varieties w.r.t $N_R(R)$
$\varepsilon_P(R)$	1.403		Elasticity of cost of consuming varieties w.r.t $N_P(R)$
$\varepsilon_R(P)$	1.321		Elasticity of cost of consuming varieties w.r.t $N_R(P)$
$\varepsilon_P(P)$	-0.772		Elasticity of cost of consuming varieties w.r.t $N_P(P)$
F	0.067		Fixed cost of consuming varieties

Note: The table summarize the estimates of parameters in section 5 using UPC-level data.

Table 9: Model Replicates Key Moments of Data

		Data	Model
Fact 1: $\frac{\partial N_r}{\partial Gini_r} > 0$		4.937 [3.614, 6.260]	5.131 [3.042, 6.538]
Fact 2: $\frac{\partial \Omega_{ri}}{\partial Gini_r} < 0$	Poor HHs	-0.504 [-0.927, -0.081]	-0.542 [-0.935, -0.111]
	Rich HHs	-0.883 [-1.498, -0.268]	-0.780 [-1.536, -0.061]

Note: The table compare the results of Fact 1 and Fact 2 in the empirical section 3 with the results obtained from model using estimated parameters. We can see that the model could replicate key data moments. The 95% confidence interval of the estimates from model is calculated using the 95% confidence interval of the estimated parameters.

Table 10: Welfare Implications of Variety Inequality

	from Low-inequality county to Median-inequality county	**from Median-inequality county to High-inequality county**
Panel A - Indirect Utility Approach (using estimated parameters)		
Poor HHs	-1.314%	-2.708%
Rich HHs	-1.523%	-7.427%
Panel B - Indirect Utility Approach (using parameters in the literature)		
Poor HHs	0.996%	1.760%
Rich HHs	0.941%	-2.454%
Panel C - Consumption Equivalence Approach (using estimated parameters)		
Poor HHs	-0.222%	-0.308%
Rich HHs	-1.110%	-2.331%

Note: The table presents the results of changes in household welfare when moving to higher-inequality counties in section 6.

A Data Appendix

The primary data sets are the Consumer Panel and Retail Scanner data. They are made available through the Kilts Center at The University of Chicago Booth School of Business. The Consumer Panel Data comprise a representative panel of 40,000–60,000 households that continually provide information about their purchases in a longitudinal study. This panel is projectable to the total United States using household projection factors. The data includes information on household demographics, geographic, product ownership, as well as select demographics for the heads of household and other members. In particular, demographic variables include household income range, size, composition, presence, and age of children, marital status, type of residence, race, and Hispanic origin. Male and female heads of households also report age range, birth date, hours employed, education, and occupation. Geographic variables include panelist zip code, FIPS state and county codes, region (East, Central, South, West), and Scantrack Market code assigned by NielsenIQ.

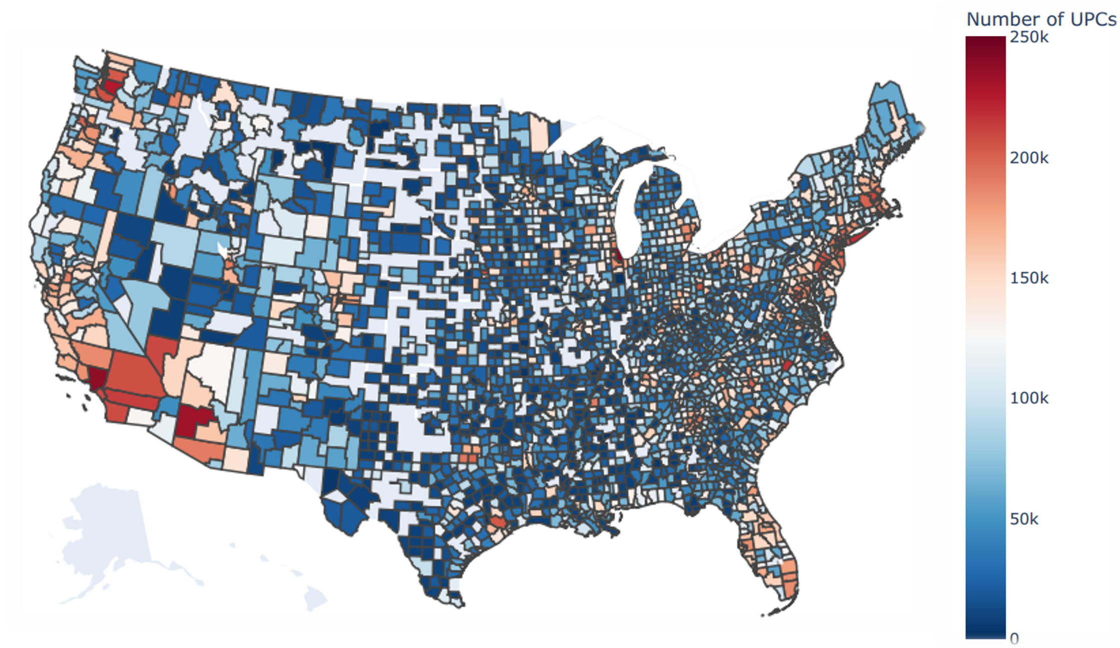
The data includes all 10 NielsenIQ food and nonfood departments (approximately 4 million UPCs). These departments are dry grocery, frozen foods, dairy, deli, packaged meat, fresh produce, nonfood grocery, alcohol, general merchandise, and health and beauty aids. Information on product characteristics, such as UPC and description, brand, multi-pack, and size, as well as NielsenIQ codes for department, product group, and product module, are available. We also observe detailed information on households' shopping trips. Each shopping trip contains the date, retail chain code, retail channel, first three digits of store zip code, and total amount spent. The UPC, quantity, price, and any deals/coupons are recorded for each product purchased.

The Retail Scanner Data consist of weekly price and quantity information generated by point-of-sale systems for roughly 90 participating retail chains across all US markets. When a retail chain agrees to share their data, all of their stores enter the database. As a result, the database includes 30,000-50,000 individual stores. The stores in the database vary widely in terms of formats and types, e.g., food, drug, mass merchandise, liquor, or convenience stores. Data entries can be linked to a store's identifier and a chain identifier, so a given store can be tracked over time and linked to a specific chain.

The two data sets have their strengths and weaknesses. The strength of the Consumer Panel Data is that it provides detailed information on household spending alongside household characteristics. Therefore, I can identify the number of UPCs consumed by a household and household expenditure share on each UPC. Its relative weakness in comparison to the Retail Scanner Data is that the Homescan sample of households only covers a fraction of the US retail market at any given period. Moreover, the data are reported by households, so it has a higher sampling error. Relative to the Homescan data, the store-level retail scanner data cover more than a thousand times the retail sales in each half year. The information is reported by the retailers, which helps to reduce measurement errors.

B Empirical Appendix

Figure 5: The number of varieties across U.S. counties



Note: The figure plots the number of products available at county level in 2011. Author's calculation using data from NielsenIQ Retail Scanner Data.

Table 11: Income Inequality and Varieties Available in County

		Number of Brands in a county					
Panel A - by County	Gini	4.004***					
		(.609)					
	N	17,893					
	R^2	0.371					
		Number of categories			Number of Brands by category		
		Department	Product Group	Product Module	by Department	by Product Group	by Product Module
Panel B - by County by Category	Gini	0.057*	0.149***	0.828***	6.089***	4.529***	3.496***
		(.021)	(.040)	(.148)	(.890)	(.647)	(.469)
	N	17,893	17,893	17,893	74,286	2,035,213	15,905,437
	R^2	0.659	0.659	0.648	0.618	0.618	0.613
		Number of stores			Number of Brands by store		
Panel C - by County by Store	Gini	6.836***			0.319**		
		(1.093)			(.145)		
	N	17,893			244,987		
	R^2	0.542			0.612		
		Number of categories by store			Number of Brands by category by store		
		Department	Product Group	Product Module	by Department	by Product Group	by Product Module
Panel D - by Store by Category	Gini	0.080**	0.139*	0.397*	.901***	0.333*	0.243*
		(.024)	(.053)	(.163)	(.162)	(.181)	(.121)
	N	244,987	244,987	244,987	2,356,515	26,464,263	175,131,402
	R^2	0.672	0.672	0.678	0.673	0.668	0.656

Note: The table presents the estimates of the specification (3.1) for the first fact — high-inequality counties have more varieties, and this results are robust along different dimensions of extensive and intensive margins. Panel A shows the most general result with dependent variable is the number of brands available in a county i in year t. Panel B disentangles the results into extensive and intensive margins by category — department, product group, product module. Panel C disentangles the results into extensive and intensive margins along the dimension of store. Lastly, Panel D shows the results for the extensive and intensive margins of the increase in the number of brands offered within a store. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 12: Extensive and Intensive margins - Varieties Purchased by Households

		Number of categories (Product Module)		Number of UPCs by category (Product Module)	
		Poor HHs	Rich HHs	Poor HHs	Rich HHs
Category	Gini	-0.450***	-0.894***	-.0143	-.237***
		(.129)	(.173)	(.073)	(.071)
	N	66,767	71,000	12,309,921	15,062,590
		Number of categories by store (Product Module)		**Number of UPCs by category by store (Product Module)**	
Category by Store		Poor HHs	Rich HHs	Poor HHs	Rich HHs
	Gini	-1.152***	-1.187***	-.126*	-.179**
		(.329)	(.279)	(.072)	(.071)
	N	418,859	466,557	15,476,447	17,762,647

Note: The table presents the estimates of the specification (3.2) for the second fact — households in high-inequality counties purchase fewer varieties, and this results are robust along different dimensions of extensive and intensive margins. The definition of category is Product Module. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by ∗ ∗ ∗, ∗∗, and ∗ respectively.

Table 13: Extensive and Intensive margins - Varieties Purchased by Households

		Number of categories (Department)		Number of UPCs by category (Department)	
		Poor HHs	Rich HHs	Poor HHs	Rich HHs
Category	Gini	0.060	0.031	-0.327**	-0.847***
		(.029)	(.029)	(.160)	(.203)
	N	66,767	71,000	621,873	685,099
		Number of categories by store (Department)		**Number of UPCs by category by store (Department)**	
Category by Store		Poor HHs	Rich HHs	Poor HHs	Rich HHs
	Gini	-.340**	-.326**	-.934***	-1.134***
		(.142)	(.132)	(.225)	(.227)
	N	418,859	466,557	2,030,854	2,269,547

Note: The table presents the estimates of the specification (3.2) for the second fact — households in high-inequality counties purchase fewer varieties, and this results are robust along different dimensions of extensive and intensive margins. The definition of category is Department. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by ∗ ∗ ∗, ∗∗, and ∗ respectively.

Table 14: Income Inequality and Brands Consumed by Households

	Number of Brands	
	Poor HHs	Rich HHs
Gini	-0.253	-0.798***
	(.159)	(.178)
R^2	0.214	0.160
N	66,767	71,00

Note: The table presents the estimates of the specification (3.2) for the second fact — households in high-inequality counties purchase fewer varieties, and this results are robust along different dimensions of extensive and intensive margins - see Table 15. The definition of variety is Brand. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 15: Extensive and Intensive Margins - Brand level

		Number of categories (Product Group)		Number of brands by category (Product Group)	
		Poor HHs	Rich HHs	Poor HHs	Rich HHs
Category	Gini	-0.275***	-0.445***	-0.269**	-0.713***
		(.077)	(.103)	(.094)	(.119)
	N	66,767	71,000	4,758,971	5,447,801
		Number of stores		**Number of brands by store**	
		Poor HHs	Rich HHs	Poor HHs	Rich HHs
Store	Gini	0.901***	0.738***	-0.867**	-0.959***
		(.201)	(.167)	(.341)	(.269)
	N	66,767	71,000	418,859	466,557
		Number of categories by store (Product Group)		**Number of brands by category by store (Product Group)**	
		Poor HHs	Rich HHs	Poor HHs	Rich HHs
Category by Store	Gini	-0.894***	-0.905***	-0.271**	-0.431***
		(.274)	(.238)	(.112)	(.113)
	N	418,859	466,557	8,479,842	9,421,181

Note: The table presents the estimates of the specification (3.2) for the second fact — households in high-inequality counties purchase fewer varieties along different dimensions of extensive and intensive margins. The definition of variety is Brand. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

Table 16: Estimation of Elasticity of Substitution σ (Brand-level Data)

	Quarterly Data			Half-year Data		
	2-quantiles	**4-quantiles**	**5-quantiles**	**2-quantiles**	**4-quantiles**	**5-quantiles**
Panel A - OLS						
$1-\sigma$	0.529***	0.490***	0.481***	0.600***	0.572***	0.564***
	(0.00573)	(0.0118)	(0.0122)	(0.00654)	(0.00819)	(0.00988)
N	8,702,123	6,022,038	4,971,010	6,132,032	4,426,200	3,725,665
Panel B – National IV						
$1-\sigma$	-1.578***	-1.532***	-1.461***	-0.609***	-0.654***	-0.615***
	(0.0479)	(0.0482)	(0.0921)	(0.0337)	(0.0554)	(0.0521)
N	8,702,123	6,022,038	4,971,010	6,132,032	4,426,200	3,725,665
First Stage F-Stat.	393.8	548.6	759.3	431.4	711.1	793.6
Panel C – State IV						
$1-\sigma$	-1.502***	-1.390***	-1.337***	-0.766***	-0.794***	-0.752***
	(0.0659)	(0.0707)	(0.0727)	(0.0489)	(0.0738)	(0.0807)
N	7,398,255	4,867,376	3,925,764	5,352,836	3,686,103	3,044,273
First Stage F-Stat.	799.6	449.4	317.3	408.7	267.3	199.1
Panel D – National and State IVs						
$1-\sigma$	-1.755***	-1.699***	-1.616***	-0.785***	-0.841***	-0.785***
	(0.0696)	(0.0797)	(0.106)	(0.0563)	(0.0711)	(0.0884)
N	7,398,255	4,867,376	3,925,764	5,352,836	3,686,103	3,044,273
First Stage F-Stat.	737.6	787.3	947.4	345.2	467.8	416.5

Note: The table presents estimates of the elasticity of substitution σ from the specification (5.1). The definition of variety is brand. 2-quantiles means splitting households by 2 groups, 4-quantiles means splitting households by 4 groups, and 5-quantiles means splitting households by 5 groups. The standard errors are clustered by county. Statistical significance at the 1%, 5%, and 10% levels are indicated by $***$, $**$, and $*$ respectively.

C Theory Appendix

C.1 Derivation of the household's indirect utility function in Equation (4.5)

First, I start by solving the household first-stage problem to obtain optimal consumption choice. Then, I show how I could rewrite the utility function in the form of equation (4.5). We set up the Lagrangian for the household's problem as:

$$\mathcal{L} = \left(\int_{\omega\in\Omega_{ri}} \left(\varphi_{i\omega} c_{ri\omega}\right)^{\frac{\sigma-1}{\sigma}} d\omega\right)^{\frac{\sigma}{\sigma-1}} - F \times (|\Omega_{\mathrm{ri}}|)^{\varepsilon} N_{rj}^{\varepsilon_j(j)} N_{r\_j}^{\varepsilon\_j(j)} - \lambda\left[\int_{\omega\in\Omega_{ri}} c_{ri\omega} p_{\omega} d\omega + C_{ri0} - E_{ri}\right]$$

where λ is the multiplier for the budget constraint. Taking the first-order condition with respect to two consumption of two different varieties $c_{ri\omega}$ and $c_{ri\omega'}$, then take the ratio of the two first-order condition, we obtain:

$$\frac{c_{ri\omega}}{c_{ri\omega\prime}} = \left(\frac{p_{\omega'}}{p_{\omega}}\right)^{\sigma} \left(\frac{\varphi_{i\omega'}}{\varphi_{i\omega}}\right)^{1-\sigma}$$

Rearrange this equation, we have:

$$c_{ri\omega} p_{\omega} = c_{ri\omega\prime} p_{\omega} \left(\frac{p_{\omega\prime}}{p_{\omega}}\right)^{\sigma} \left(\frac{\varphi_{i\omega\prime}}{\varphi_{i\omega}}\right)^{1-\sigma}$$

Then we take integral across all varieties $\omega \in \Omega_{ri}$:

$$E_{ri} - C_{ri0} = \int_{\omega\in\Omega_{ri}} c_{ri\omega} p_{\omega} d\omega = c_{ri\omega\prime} p_{\omega'}^{\sigma} \varphi_{i\omega'}^{1-\sigma} \int_{\omega\in\Omega_{r,i}} \left(\frac{\varphi_{i\omega}}{p_{\omega}}\right)^{\sigma-1} d\omega$$

We define the price index as $P_{ri} = \left[\int_{\omega\in\Omega_{ri}} \left(\frac{\varphi_{i\omega}}{p_{\omega}}\right)^{\sigma-1} d\omega\right]^{\frac{1}{1-\sigma}}$, then we can obtain the optimal consumption choice of the household as in the equation (4.2). We can also rewrite the equation above as:

$$c_{ri\omega}\varphi_{i\omega} = \left(E_{ri} - C_{ri0}\right) P_{ri}^{\sigma-1} \left(\frac{\varphi_{i\omega}}{p_{\omega}}\right)^{\sigma}$$

We raise both sides to the power $\frac{\sigma-1}{\sigma}$ and take integral across all varieties $\omega \in \Omega_{ri}$:

$$\int_{\omega\in\Omega_{ri}} \left(\varphi_{i\omega} c_{ri\omega}\right)^{\frac{\sigma-1}{\sigma}} d\omega = \left(E_{ri} - C_{ri0}\right)^{\frac{\sigma-1}{\sigma}} P_{ri}^{\frac{(\sigma-1)^2}{\sigma}} \int_{\omega\in\Omega_{ri}} \left(\frac{\varphi_{i\omega}}{p_{\omega}}\right)^{\sigma-1} d\omega = \left(E_{ri} - C_{ri0}\right)^{\frac{\sigma-1}{\sigma}} P_{ri}^{\frac{(\sigma-1)^2}{\sigma}} P_{ri}^{1-\sigma}$$

Lastly, we raise both sides to the power $\frac{\sigma}{\sigma-1}$ to obtain the equation:

$$\left[\int_{\omega\in\Omega_{ri}} \left(\varphi_\omega c_{ri\omega}\right)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}} = \frac{E_{ri} - C_{ri0}}{P_{ri}}$$

C.2 Derivation of the price index in Equation (4.8)

To derive the final equation of the price index in the equation (4.8), we start with the definition of the price index in the equation (4.3):

$$\begin{aligned}
P_{ri} &= \left[\int_{\omega\in\Omega_{ri}} \left(\frac{\varphi_\omega}{p_\omega}\right)^{\sigma-1} d\omega\right]^{\frac{1}{1-\sigma}} = N_r^{\frac{1}{1-\sigma}} p \left[\int_{\bar{\varphi}^*_{ri}}^{\infty} y^{\sigma-1} \frac{\theta\varphi_m^\theta}{y^{\theta+1}} dy\right]^{\frac{1}{1-\sigma}} \\
&= N_r^{\frac{1}{1-\sigma}} p \varphi_m^{\frac{\theta}{1-\sigma}} \theta^{\frac{1}{1-\sigma}} \left[\int_{\bar{\varphi}^*_{ri}}^{\infty} y^{\sigma-\theta-2} dy\right]^{\frac{1}{1-\sigma}} \\
&= N_r^{\frac{1}{1-\sigma}} p \varphi_m^{\frac{\theta}{1-\sigma}} \theta^{\frac{1}{1-\sigma}} \left[\frac{1}{\sigma-\theta-1}\right]^{\frac{1}{1-\sigma}} y^{\frac{\sigma-\theta-1}{1-\sigma}} \Big|_{\bar{\varphi}^*_{ri}}^{\infty}
\end{aligned}$$

Using the assumption of Pareto-distributed taste parameter which gives us the equation (4.7), we can obtain the final equation for the index price as in equation (4.8).

C.3 Derivation of firm's optimal quality in Equation (4.11)

Recall firm's profit:

$$\Pi_\omega = R_{r\omega}\left(1 - \frac{1}{M}\right) - b\cdot\beta\cdot\phi^{\frac{1}{\beta}} - f c_{0\omega} - f$$

where $M = \frac{\sigma}{\sigma-1}$ is firm's markup. Then take FOC w.r.t $log(\phi)$:

$$0 = \frac{\partial\Pi}{\partial\log(\phi)} = \frac{\partial\Pi}{\partial(\phi)}\cdot\frac{\partial\phi}{\partial\log(\phi)} = \phi\frac{\partial\Pi}{\partial(\phi)}$$
$$= \phi\left[\left(1-\frac{1}{M}\right)\int_j \frac{\partial R_{rj\omega}}{\partial\phi} dj - b\phi^{\frac{1}{\beta}-1}\right] = \left(1-\frac{1}{M}\right)\int_j (\sigma-1)\left(\gamma_j - \xi\right) R_{rj\omega} dj - b\phi^{\frac{1}{\beta}}$$

Recall from FOC of firm's third decision:

$$\left(1-\frac{1}{M}\right) = \frac{1}{M}\frac{\int_j R_{rj\omega} dj}{\int_j (\sigma-1) R_{rj\omega} dj}$$

Thus firm's optimal quality is

$$
\begin{aligned}
\phi &= \left[\frac{1}{b}\left(1-\frac{1}{M}\right)\int_j (\sigma-1)\left(\gamma_j-\xi\right) R_{rj\omega} dj\right]^{\beta} \\
&= \left[\frac{1}{bM}\left(\frac{\int_j \gamma_j(\sigma-1)R_{rj\omega}dj}{\int_j^1 (\sigma-1)R_{rj\omega}dj}-\xi\right)\int_j R_{rj\omega}dj\right]^{\beta} \\
&= \left[\frac{1}{bM}(\tilde{\gamma}-\xi)R_{r\omega}\right]^{\beta}
\end{aligned}
$$

C.4 Proof: $\frac{\partial \delta}{\partial \mathbf{Gini}} > 0$

Recall that Gini coefficient of a continuous distribution with mean μ:

$$
\text{Gini} = 1 - \frac{1}{\mu}\int_0^{+\infty}(1-\text{F(y)})^2 dy = \frac{1}{\mu}\int_0^{+\infty} F(y)(1-\text{F(y)})dy
$$

If the distribution is log-normal(μ, δ^2), then the Gini index is:

$$
\text{Gini } = 2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1
$$

We know a normal distribution $N\left(\mu, \delta^2\right)$ has CDF $F_X\left(x, \mu, \delta^2\right) = \Phi\left(\frac{x-\mu}{\delta}\right)$. Then

$$
\frac{\partial}{\partial \delta}F_X\left(x, \mu, \delta^2\right) = \frac{\partial}{\partial \delta}\Phi\left(\frac{x-\mu}{\delta}\right) = -\left(\frac{x-\mu}{\delta^2}\right)\phi\left(\frac{x-\mu}{\delta}\right)
$$

Thus

$$
\frac{\partial gini}{\partial \delta} = \frac{\partial}{\partial \delta}\left[2\Phi\left(\frac{\delta}{\sqrt{2}}\right)-1\right] = \frac{1}{\sqrt{2}}2\phi\left(\frac{\delta}{\sqrt{2}}\right) = \sqrt{2}\phi\left(\frac{\delta}{\sqrt{2}}\right) > 0
$$

Then

$$
\frac{\partial \delta}{\partial \,\text{Gini}} = \frac{1}{\sqrt{2}\phi\left(\frac{\delta}{\sqrt{2}}\right)} > 0
$$

C.5 Proof of Proposition 2

Notice that we have

$$
\frac{\partial \log\left(N_{rR}\right)}{\partial \log\left(N_r\right)} = \frac{\partial \log\left(N_r - N_{rP}\right)}{\partial N_r} \cdot \frac{\partial N_r}{\partial \log\left(N_r\right)} = \frac{N_r}{N_r - N_{rP}} = \frac{N_r}{N_{rR}} \quad \text{and similarly, } \frac{\partial \log\left(N_{rP}\right)}{\partial \log\left(N_r\right)} = \frac{N_r}{N_{rP}}
$$

And,

$$
\frac{\partial \log\left(N_r\right)}{\partial \log\left(N_{rR}\right)} = \frac{\partial \log\left(N_{rR}+N_{rP}\right)}{\partial N_{rR}} \cdot \frac{\partial N_{rR}}{\partial \log\left(N_{rR}\right)} = \frac{N_{rR}}{N_{rR}+N_{rP}} = \frac{N_{rR}}{N_r} \quad \text{and similarly, } \frac{\partial \log\left(N_r\right)}{\partial \log\left(N_{rP}\right)} = \frac{N_{rP}}{N_r}
$$

And,

$$\frac{\partial \log(N_{rP})}{\partial \log(N_{rR})} = \frac{\partial \log(N_r - N_{rR})}{\partial N_{rR}} \cdot \frac{\partial N_{rR}}{\partial \log(N_{rR})} = \frac{-N_{rR}}{N_r - N_{rR}} = -\frac{N_{rR}}{N_{rP}} \text{ and similarly, } \frac{\partial \log(N_{rR})}{\partial \log(N_{rP})} = -\frac{N_{rP}}{N_{rR}}$$

Then from the equation (4.9) of the solution of the optimal number of varieties household consume, we obtain:

$$\frac{\partial \log(|\Omega_{ri}|)}{\partial \log(N_r)} = \frac{1}{\varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}} \left[\frac{1}{\theta} - \varepsilon_j(j) \frac{N_r}{N_{rj}} - \varepsilon_{-j}(j) \frac{N_r}{N_{r_-}} \right] < 0$$

Under the conditions for parameters in Proposition 2, we obtain $\frac{\partial \log(|\Omega_{ri}|)}{\partial \log(N_r)} < 0$

C.6 Derivation of Equation (5.3)

First we start with the solution of household's first-stage problem and take variance of both sides.

$$\mathrm{Var}(|\Omega_{ri}|) = \left[\frac{\varphi_m}{\varepsilon F} \left(\frac{1}{\sigma - 1} - \frac{1}{\theta} \right) \left(1 + \frac{1-\sigma}{\theta} \right)^{\frac{1}{1-\sigma}} N_r^{\frac{1}{\theta}} N_{rj}^{-\varepsilon_j(j)} N_{r_j}^{-\varepsilon_j(j)} \right]^{\frac{2}{\varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}}} \times \mathrm{Var}\left((E_{ri} - C_{ri0})^{\frac{1}{\varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}}} \right)$$

Then we divide both sides by $(|\Omega_{ri}|)^2$ as the number of varieties consumed by households is positive.

$$\frac{\mathrm{Var}(|\Omega_{ri}|)}{(|\Omega_{ri}|)^2} = \frac{\mathrm{Var}\left((E_{ri} - C_{ri0})^{\frac{1}{\varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}}} \right)}{(E_{ri} - C_{ri0})^{\frac{2}{\varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}}}}$$

Finally we take log of both sides to obtain the equation (5.3):

$$\log\left(\frac{\mathrm{Var}(|\Omega_{ri}|)}{(|\Omega_{ri}|)^2} \right) = \log\left(\mathrm{Var}\left((E_{ri} - C_{ri0})^{\frac{1}{z}} \right) \right) - \frac{2}{z} \log(E_{ri} - C_{ri0})$$

where $z = \varepsilon + \frac{1}{\theta} + \frac{1}{1-\sigma}$