Effects of Income Distribution Changes on Assortment Size in the Mainstream Grocery Channel

Rafael Becerril Arreola

USC Darla Moore School of Business

Randolph E. Bucklin

UCLA Anderson School of Management

Raphael Thomadsen

Olin Business School, Washington University in St. Louis

ABSTRACT

The authors study the effect of changes in the U.S. income distribution on assortment size in the mainstream grocery channel. Census demographics for 1,711 counties are matched to local assortment data from Nielsen in 944 grocery product categories from 2007 to 2013. The authors show that—holding other demographics constant—assortment size is increasing with average income but decreasing with the Gini index of income. We demonstrate that most of this effect occurs through the changes in customer spending that would be predicted by Engel's law for necessity-type goods. Even after controlling for spending we continue to find a negative effect of the Gini index on assortment, but at a reduced level. We examine factors that may explain this finding. The effects of income on assortment size occur for both horizontally and vertically differentiated dimensions of assortment, with a larger effect for horizontal differentiation. The effect sizes are similar for private labels and branded products. The findings offer CPG manufacturers insights into long-term market trends and which markets should receive resources to expand shelf presence versus defend a current position.

Keywords: Income distribution, Census data, consumer packaged goods, retailing, assortment decisions

1. Introduction

National and international policymakers, economists, management consultants, and the media have recently drawn increased attention to changes in the distribution of income over the past several decades (e.g., Dabla-Norris et al. 2015; Derby 2015). Census Bureau data show that inflation-adjusted average income has oscillated widely: For example, median income dropped nine percent from a high in 2007 to 2012, before recovering to its 2007 level in 2016. Income dispersion has risen significantly in the U.S., with the Gini index moving from .40 in 1980 to .48 in 2017, a 20 percent increase. Figure 1 shows Census data for incomes at various percentiles from 1967 to 2017; incomes of top earning households expanded dramatically, while those of the lower tiers have remained relatively flat. Such shifts in the distribution of income can lead to changes in aggregate consumer spending (e.g., Krueger and Perri 2006; Kamakura 2014) with implications for firms and marketing strategies.

Industry experts have expressed conflicting views about how firms should respond to changes in the income distribution, with some advocating that companies target the pressing needs of low-income consumers (e.g., Simanis and Duke 2014) and others advising that high-income consumers be targeted with premium products (e.g., Lodd 2016). The high-end grocery retailer Whole Foods entered low-income markets seeking to find new ways to sustain growth (Berman 2014), whereas many mass-market retailers tried to attract high-income consumers with upscale merchandise and nicer store layouts (Ries 2013).

The issue of how to react to changes in income and spending patterns is especially important for Consumer Packaged Goods (CPG) manufacturers, who have to continuously fight for shelf space. Retailers frequently add and subtract products from their assortments. Manufacturers need to understand what factors drive these shifts in the retail channel so resources can be allocated to either defend their territory in contracting markets or seek new shelf space in expanding markets. It is also important for manufacturers to understand which products are most affected by changes in incomes.

In this paper we focus on the effect that changes in the distribution of income have on assortment in the mainstream grocery channel. As the income distribution changes—and with that, spending in the food retailing industry—retailers may both alter the amount of space they allocate to CPG products (compared

to deli, bakery, meats, and other non-CPG categories), as well as change the composition of the products they offer within each category. Further, as the income distribution changes, the incentives for retailers to match versus differentiate from competitors through assortment can shift. Thus, the total assortment that is available to customers—within each store but also across stores within a market—is likely to vary with the income distribution.

We examine how the size of the assortment of products available to consumers changes with average income and the Gini index—the classic measure of income dispersion or income inequality (terms we will use interchangeably). We also consider which mechanisms are most important in explaining these effects. We then consider what types of products are most affected by shifts in the income distribution.

The question of how income distribution changes affect assortment is theoretically ambiguous and therefore an empirical question. Our review of the literature suggests that there are at least three reasons why income can affect assortment and we discuss these in turn.

First, Engel's Law for variety documents that increasing incomes lead to higher within-household tastes for consumption variety (Chai, Rohde, and Silber 2015; Carlson et al. 2015); thus, higher average income may be expected to lead retailers to opt for larger assortments. Greater income dispersion also may lead to an increase in assortment size if most of the increasing taste for variety comes from upper tiers of the income distribution.

Second, theoretical models of vertical differentiation (e.g., Gabszewicz and Thisse 1979; Shaked and Sutton 1983) suggest that income dispersion could lead to a broader offering of products. The underlying assumption behind the vertical differentiation theories is that greater income dispersion is equivalent to greater consumer heterogeneity. One caveat for this effect is that more income inequality could arise from a "hollowing out" of middle-income consumers, leaving a greater concentration of lower-income and upper-income households. This shift towards the extremes would make the population more homogenous on the low end as income inequality increases, leading overall consumer heterogeneity to actually decline as income inequality rises. This could produce the counterintuitive result that greater inequality leads to smaller assortments.

Finally, both average income and income inequality can affect assortment size through their effects on spending in the supermarket industry. Higher average incomes can increase spending (Lancaster 1990). Further, if grocery categories are, for the most part, necessities, spending will rise less than proportionally with income (Engel 1857; Du and Kamakura 2008; Lewbel 2008; Cirera and Masset 2010), producing a concave relationship between household income and spending on groceries that is known as Engel's Law for expenditure. At the aggregate level, the concavity implies that an increase in income dispersion—holding average income constant—decreases the overall category expenditure, because high-income households will not increase their spending as much as low-income households decrease theirs, as shown in a stylized fashion in the top panel of Figure 2. The lower panel of Figure 2 shows the actual income and expenditure levels of the households in the Nielsen panel-data sample, along with the fitted mean relationship, which is concave. A decrease in spending that accompanies increased income inequality could lead firms to reduce their assortments because retailers cannot support the fixed costs of stocking some varieties of products when total spending decreases.

Drawing upon Nielsen and the U.S. Census Bureau as sources, we assemble a large-scale dataset that describes both demographic characteristics and market outcomes in 944 CPG product categories across 1,711 U.S. counties over a period of seven years (2007-2013). The national scale of the data enables us to study the phenomenon in the U.S. as a whole. An important feature of the data is that U.S. counties varied substantially in the income distribution changes they experienced: Though the distribution became more unequal in the majority of counties, income dispersion actually fell in a large proportion of counties. We measure changes in the county-level income distributions with the Gini index, as reported by the U.S. Census Bureau, which is widely used to capture the inequality or dispersion of the income distribution (e.g., Cowell and Flachaire 2013; Leigh 2007). A value close to zero represents equal incomes, whereas a Gini close to one indicates that nearly all income is received by one household. We measure assortment size using the number of UPCs available, which is well suited for the large-scale analysis we conduct and has

been widely used in the grocery industry to track and manage assortments (e.g., Salazar 2014; Watson 2014; Draganska, Klapper, and Villas-Boas 2010; Hwang et al. 2010; Ren et al. 2011; Jaravel 2018).[1]

We find that increased average income and decreased income dispersion lead to larger assortments and that this result is robust to a recession: the effects of change in income on changes in assortment are similar for 2007-2010 and 2010-2013. We then explore which of the three mechanisms we just discussed explain this effect. We show that most of this effect can be explained by greater spending that occurs when average income increases and income dispersion decreases. However, we find that the negative relationship between assortment and the Gini index persists even after controlling for spending. We offer some evidence that this may be because of the hollowing out of the middle class with increased income inequality.

We then examine which products are most affected by changes in income. We find that the changes in average income and inequality affect both vertical and horizontal differentiation in the same direction, but the effect is larger for horizontal differentiation. The impact of changes in income inequality on UPC counts is similar for private labels and national brands, but higher Gini indexes lead to large brands losing a greater fraction of their UPCs than small brands.

Our paper is related to research that studies how retailers—or retailers and manufacturers jointly—rely on local consumer demographics to plan their assortments (Grewal et al. 1999; Kadiyali, Chintagunta, and Vilcassim 2000; Israilevich 2004; Rosenblum and Rowen 2008; Dukes, Geylani, and Srinivasan 2009). For example, Hwang et al. (2010) show that high-end regional brands of cereal and toothpaste were more likely to be carried by retailers in areas with high proportions of high-income households, demonstrating an empirical link between income and retail assortment decisions. As in their study, we do not attempt to formally model retailer or manufacturer behaviors.

Our topic is also related to the literature on the "poverty penalty," which examines how low-income consumers may face higher food prices (e.g., Kunreuther 1973; Talukdar 2008) and may reside in areas with limited access to groceries (so-called "food deserts"). More recently, Orhun and Palazzolo (2019) find

[1] We also note that more complex measures designed to capture the variety in an assortment are often highly correlated with the count-based measure of assortment size (see, e.g., Van Herpen and Pieters 2002).

that liquidity constraints can inhibit the ability of low-income households to buy large packages of non-perishable goods, so they are less likely to benefit from quantity discounts. Allcott, Diamond, and Dubé (2017) look at how the mix of stores affects the nutrition of consumers in poorer neighborhoods. They note that different stores offer wider or narrower selections, and that poorer areas have more stores with limited selection (e.g., dollar stores). Allcott et al. then use changes in the presence of stores to measure whether the supply of different UPCs affects the nutritional value of the choices by poorer consumers, and find that supply has only a small effect. Our research question differs in that we focus on how assortment changes with shifts in the overall distribution of income over time. Allcott et al. does raise another potential mechanism for our result, which is that the mix of stores could change as income inequality changes. Although such a shift could explain some of our results, we find that the effect of income inequality on product assortment also occurs within-stores, and that the majority of the effect can be explained even at this level. Also, much of the previous literature has focused on only what happens with the lowest income brackets, while we are concerned with the assortment of products offered to consumers in all income brackets.

The balance of this paper is structured as follows. Section 2 provides a detailed description of our data. Section 3 presents our model and empirical strategy. Our main findings are in Section 4, which also presents robustness checks and explores the underlying mechanisms. Section 5 explores the managerial implications of this relationship, and Section 6 concludes.

2. Data

Most of the income and demographic information we use comes from the U.S. Census Bureau. Though the decennial census covers the entire U.S. population, it does not provide the longitudinal data we need because it is conducted only once every 10 years. The American Community Survey (ACS) samples 3 million addresses in the U.S. every year, and the Census Bureau uses the survey to create population estimates. These population estimates are available at the levels of county, county subdivision, place, congressional district, school district, and larger areas. Up to 2013, the Census Bureau also provided annual

estimates that pool data for 3-year and 5-year rolling periods to reduce sampling error. The pooled estimates are based on a larger number of respondents per geographic area and also are reported for a much larger number of counties than the individual-year estimates. We use the annually reported 3-year rolling-sample estimates and select the county as the geographic unit of analysis. In total, we compile data for 1,711 counties that consistently appear in the 3-year estimates of the American Community Survey from 2007 to 2013. These counties correspond to 204 Designated Market Areas (DMAs) and account for 94% of the U.S. population in the 48 contiguous states.[2] The variables used in the main analysis are described in Web Appendix W1. Descriptive statistics and correlations appear in Tables 1 and 2. Although an effort could be made to work with data at the county subdivision level to increase granularity, the sample would be limited because the Census does not report subdivision data for many areas.

Our dependent variables are derived from the Nielsen Retail Scanner dataset covering the 2007-2013 period, as distributed by the Kilts Center for Marketing. The dataset reports weekly unit sales and average prices from about 35,000 stores belonging to approximately 90 participating retail chains operating in the 48 contiguous states.[3] The location of each store is reported in terms of its Federal Information Processing Standard (FIPS) county and state codes, DMA code, and the first three digits of its ZIP code, but the identity of the store is not revealed. The dataset includes approximately 2.6 million UPCs within 1,075 product modules (categories) in 10 departments. The all-commodity volume (ACV) by channel is 53% of grocery stores, 55% of drugstores, 32% of mass merchandisers, 1% of liquor stores, and 2% of convenience stores.

We use the following departments for analysis: Frozen Foods, Health and Beauty Care, Dry Grocery, Non-Food Grocery, and General Merchandise. Other departments for which data were available include

[2] The 3-year estimates of the American Community Survey are reported for a total of 1,846 (59%) of the 3,141 counties and county equivalents in the U.S. (We select 1,711 on the basis of consistent reporting.) These estimates correspond to geographic areas with at least 20,000 inhabitants. We examined the differences between the 1,846 included counties and the 1,295 excluded ones using data from the year 2000 decennial census. With the exception of population size, we found that the two sets of counties differ only slightly in terms of their demographics, such as average income. Unfortunately, even the decennial census does not report income dispersion in small-population counties, so assessing how they might differ from the counties we analyze in this respect is not possible.

[3] Nielsen does not report data from all affiliated retailers, and on rare occasions, a few stores may be excluded from the sample because of confidentiality concerns.

Alcohol, Dairy, Deli, Packaged Meat, and Fresh Produce. We exclude Alcohol because the industry is highly regulated at the state level. We do not include perishables because their assortment is strongly affected by difficult-to-control factors such as the quality of distribution-related infrastructure and the distance between retailers and producers. We also drop categories that do not appear to be available consistently in each county every year. Altogether, these steps yielded a total of 944 product categories for analysis. To match with the census data at the county level, we aggregate the product-category information across stores within each county and within each year.

We next illustrate the variation in the data that allows us to empirically identify the effect of changes in income distributions on changes in product assortment. The top panel of Figure 3 presents the mean, across counties, of the county-level real average income, Gini index, and average (across categories) number of UPCs for each year in the sample, together with their 95% confidence intervals, which shows that real incomes have fallen and average inequality has significantly increased over time. The bottom panel of Figure 3 presents histograms of the changes of the same variables between 2007 and 2013. While average income decreased and income inequality increased in the U.S. overall, significant variation exists at the county level, and average income actually rose and income inequality actually fell for a significant number of counties. As is the case for average income and the income Gini, we note that the changes in county-level assortments also vary significantly.

The left panel of Figure 4 depicts the changes in UPCs between 2007 and 2013. We plot these changes separately for counties with decreased average income and those with increased average income over the same time span. For counties with increased average income, the gains in product assortment are significantly higher than the gains in counties with decreased average income ($p < .01$). The right panel of Figure 4 depicts the changes in UPC per $10,000 of local average income between 2007 and 2013; that is, we divide the local number of UPCs by the local average income, expressed in units of 10,000 dollars.[4] We

[4] The division by average income is done for graphic purposes as a way to control for average income, because the relationship between income inequality and spending (and therefore UPC availability) is only hypothesized to exist when average income is held constant. However, as we show in Table 2, higher Gini indices are correlated with lower average incomes.

compute these changes separately for counties with decreased and increased Gini index values. For counties with an increase in the Gini, the gains in product assortment are significantly lower than the gains in counties with a decrease in the Gini ($p < .05$). We also ran a Fisher Exact Test, which reject the hypotheses that changes in product assortment are independent of changes in average income and of changes in the Gini index ($p < .001$). This model-free evidence suggests that an increase in average income will be associated with an enlargement of grocery product assortment, while an increase in the Gini index will be associated with a reduction.

3. Empirical Model and Estimation

Model and Identification

We now introduce our econometric approach to estimate the relationship between changes in the income distribution and changes in product assortment. In contrast with recent work that has used cross-sectional regressions to study the effects of demographics on product assortment (Hwang et al. 2010) and product availability (Handbury and Weinstein 2014), we use panel regressions to explain the geographic variation of changes in product assortment in response to changes in the income distribution. Our analysis thus focuses on variation that is both cross-sectional and longitudinal, effectively controlling for variation that is purely longitudinal (e.g., the 2008 recession) or purely cross-sectional (e.g., climate differences and constant supply factors such as distance to producers). In our approach, we employ a combination of fixed-effects regression and feasible generalized least squares (FGLS).

Because income distribution values are specific to each county and year combination, we regard counties as the main observational units and changes in average income and Gini as heterogeneous treatments that can vary across units. We capitalize on the three-dimensional nature of our dataset to improve statistical power by treating product categories as repeated observations for each county and year combination while controlling for category-specific intercepts and trends. We identify the effects of changes in the income distribution on product assortment by exploiting the heterogeneous variation in the

direction and magnitude of the annual *changes* in the average income, the Gini index and the product assortment available, as discussed above.

Let $nUPCs_{ict}$ denote the number of distinct UPCs offered in category *i* by stores in county *c* and year *t*. We express this measure of assortment as a linear function[5] of average income ($AVGINCOME_{ct}$), the Gini index ($GINI_{ct}$), and control variables contained in the matrix X_{ct}. We employ control variables that exhibit both longitudinal and cross-sectional variation, and are known to determine retailers' assortment decisions. In addition, we include controls that have been shown to influence product availability including population size and proxies for the cost of living and the diversity of consumer preferences (Handbury and Weinstein 2014). We also control for the number of retail chains and stores in each county to account for any possible change in the set of stores and retailers that report to Nielsen. We include fixed effects specific to each combination of category and time period, τ_{it}, to account both for category-specific time trends such as category life cycles and for manufacturer supply considerations, such as manufacturing costs and factor inputs, which are largely national in nature. In addition, we include category-county specific fixed effects, γ_{ic}, to account for geographic differences in consumer preferences for specific categories (e.g., demand for mosquito repellent may be stronger in Alabama counties than in California counties) and for retailer considerations, such as distribution costs. Thus, we estimate

$$\log(nUPCs_{ict}) = \beta_0 AVGINCOME_{ct} + \beta_1 GINI_{ct} + \tau_{it} + \gamma_{ic} + \varepsilon_{ct} + X_{ct}\boldsymbol{\omega} + u_{ict}, \tag{4}$$

where ε_{ct} and u_{ict} are error terms assumed to satisfy E[ε_{ct}]=0 and E[u_{ict}|θ]=0, θ={$\beta,\tau,\gamma,\boldsymbol{\omega}$}.[6]

The coefficients of main interest are β_0 and β_1, which measure the magnitude of the effects that local changes in average income and income Gini have on local assortments in *equilibrium*. Changes in the income distribution may influence assortments directly through changes in consumer demand but also

[5] Although the linear functional form can be derived from aggregating Engel curves, we also ran some analyses with nonlinear functions as a robustness check. See Web Appendix W2.

[6] Note that full statistical independence between u_{ict} and x_{ct} is not required to identify average treatment effects (Athey and Imbens 2006). A sufficient condition for E[u_{ict}|θ]=0 to hold is that u_{ict} be statistically independent of group and time (Santos Silva and Tenreyro 2006). This condition implies that VAR($nUPCs_{ict}$|θ) is a function of the variance of u_{ict}; that is, income inequality and the controls may shift not only the conditional mean but also the conditional variance. To account for potential deviations from this assumed correlational structure, we use estimation methods robust to heteroscedasticity and serial correlation.

indirectly because retailers and manufacturers may respond to those changes in demand by adjusting other elements of the marketing mix (e.g., prices and store format) and these adjustments may further affect demand and assortments. We do not seek to explain these other changes in the marketing mix. Rather, we focus purely on understanding how the income distribution changes the equilibrium total assortment offered in the market.

To estimate our model, we first note that all fixed effects τ_{it} and γ_{ic} can be removed with the covariance estimator (Hsiao 2014, p. 62), which transforms (4) into

(5) $$\Delta\Delta y_{ict} = \beta_0 \Delta\Delta AVGINCOME_{ct} + \beta_1 \Delta\Delta GINI_{ct} + \Delta\Delta\varepsilon_{ct} + \Delta\Delta X_{ct}\boldsymbol{\omega} + \Delta\Delta u_{ict},$$

where

$$\Delta\Delta y_{ict} = \log(nUPCs_{ict}) - \overline{\log(nUPCs_{i\cdot t})} - \overline{\log(nUPCs_{ic\cdot})} + \overline{\overline{\log(nUPCs_{i\cdot\cdot})}},$$

$$\Delta\Delta AVGINCOME_{ct} = AVGINCOME_{ct} - \overline{AVGINCOME_{\cdot t}} - \overline{AVGINCOME_{c\cdot}} + \overline{\overline{AVGINCOME_{\cdot\cdot}}},$$

$$\Delta\Delta GINI_{ct} = GINI_{ct} - \overline{GINI_{\cdot t}} - \overline{GINI_{c\cdot}} + \overline{\overline{GINI_{\cdot\cdot}}},$$

$$\Delta\Delta\varepsilon_{ct} = \varepsilon_{ct} - \bar{\varepsilon}_{\cdot t} - \bar{\varepsilon}_{c\cdot} + \bar{\bar{\varepsilon}}_{\cdot\cdot},$$

$$\Delta\Delta X_{ct} = X_{ct} - \overline{X_{\cdot t}} - \overline{X_{c\cdot}} + \overline{\overline{X_{\cdot\cdot}}}$$

$$\Delta\Delta u_{ict} = u_{ict} - \bar{u}_{i\cdot t} - \bar{u}_{ic\cdot} - \bar{\bar{u}}_{i\cdot\cdot}.$$

Note that a single bar denotes an average along one dimension and double bars denote an average along two dimensions, and that the averages for X are calculated column-by-column. The dot in the subindex denotes the dimension along which the averaging takes place.

Because we consider a wide set of product categories with very different characteristics such as sales volumes, we expect that $\Delta\Delta u_{ict}$ and $\Delta\Delta\varepsilon_{ct}$ might exhibit heteroscedasticity. In addition, the unobservables ε_{ct} might be serially correlated, which would translate to serial correlation in $\Delta\Delta\varepsilon_{ct}$. Finally, different retailers serve different regions. Thus, their actions might lead to spatially correlated residuals (Bronnenberg and Mahajan 2001). Spatially correlated residuals may also arise from consumers shopping at stores located in counties that are different from the ones in which they live. Heteroscedasticity, serial correlation, and spatial correlation may induce biases in the estimates of the standard errors (Barrios,

Diamond, Imbens, and Kolesár 2012). Thus, we utilize an FGLS approach. We estimate different specifications of the correlation of the residuals, including (1) no correlation, (2) spatially correlated errors, (3) a combination of cross-sectional and serial correlation, and (4) a combination of cross-sectional, serial, and spatial correlation.

We note that our model can account for several sources of omitted variables that could explain product assortment and might be correlated with the distribution of income. For instance, some aspects of the distribution of unobserved consumer preferences could be omitted, correlate with average income or the Gini index, and determine assortment. As noted by Dubé, Hirsch, and Rossi (2018), households with stronger preferences for branded products could self-select into careers associated with high income. Likewise, households with weak preferences for CPGs and strong preferences for prepared meals could self-select into high-paying careers. However, these kinds of self-selection are controlled for by the county-level fixed effects (Dubé, Hirsch, and Rossi 2018). These fixed effects will control for all fixed parts of the consumers' preferences so that, to create biases in our analyses, unobserved aspects of preferences would have to be the part of the preferences that evolve over time. In a study that tracked the behavior of consumers who relocated across state boundaries, Bronnenberg, Dubé, and Gentzkow (2012) showed that, even after large changes in supply-side factors (operationalized in their study as the relocations), preferences for favored brands change very slowly. Hence, we expect that, over the seven years we study, the observed changes in product assortment will be more heavily influenced by changes in income than by changes in innate preferences. Further, we obtain similar results when we estimate our model on shorter time periods, as reported in Section 4. We also note that we control for category-specific trends and county-level preferences for categories through the fixed effects τ_{it} and γ_{ic} (both of which are differenced out instead of estimated).

The income distribution could also be correlated with unobserved supply factors that determine retailer and manufacturer decisions to offer certain products. As with consumer preferences, retailer self-selection and other stable factors are absorbed into county fixed effects. Yet, because our identification strategy relies on variation that is both cross-sectional and longitudinal, our estimates are still susceptible

to contamination from changes in supply factors that are both year and county specific. For example, local producers and retailers could open or close down their establishments, simultaneously affecting the local availability of products and the incomes of the local population. We mitigate potential biases that could arise from ignoring these unobserved factors in two ways. First, we exclude from our analysis the types of categories where local changes in supply factors and regulations might have a strong effect on available assortment (e.g., highly perishable products and alcohol).[7] Second, we include county-time and category-time fixed effects, along with a spatially correlated error structure (Bronnenberg and Mahajan 2001), to control for unobserved national and regional supply shocks. Unobserved retailer- and store-specific factors are discussed below together with the robustness of the results.

Estimation

Researchers have proposed iterative projection methods to estimate linear models with clustered errors on datasets similar to ours (e.g., Correia 2017), but these methods have not yet been extended to account for serially or spatially correlated errors. Because allowing errors to be clustered does not address the problem of unobserved spatial correlation (Barrios et al. 2012), FGLS methods are needed. The implementation of FGLS, however, involves computing the disturbance covariance matrix $\Omega = u_{ict} u_{ict}^T$. Because our sample includes 11,306,288 observations, storing the 1×10^{13} elements of this matrix would require 354TB of memory (in single precision). Thus, given the large scale of our study, directly implementing FGLS for the proposed model in (2) is not computationally possible. To address this challenge, we turn to a two-step procedure described by Hansen (2007). The first step of this procedure involves using ordinary least squares (OLS) to estimate the fixed-effects regression

(6) $$y_{ict} = \delta_{ct} + \tau_{it} + u_{ict}.$$

The second step involves using robust methods to regress the fixed effects δ_{ct} on the covariates of interest. When the number of categories is large, as is the case in our application with 944 categories, the

[7] An alternative approach to accounting for supply-side or cost factors that are likely to affect manufacturer and retailer decisions would be to model these decisions explicitly. Given the national scale and the local emphasis of our study, such a modeling effort is beyond the scope of this paper.

estimates $\hat{\delta}_{ct}$ can be treated as data. The estimation of the first stage with OLS and the second stage with robust methods then yields consistent estimates in the face of clustering, serial correlation, and uncertainty in the estimation of $\hat{\delta}_{ct}$ (Hansen 2007). The estimates of β_0, β_1, and $\boldsymbol{\omega}$ are unbiased and consistent even if $\hat{\delta}_{ct}$ are estimates rather than data (Hansen 2007). By following this two-step procedure, we reduce the computational scale of the estimation problem from 11,306,288 observations to only 11,977, so we can accommodate a non-i.i.d. error structure in the second step.

The estimation of equation (6) also involves challenges. Our model includes 11,977 fixed effects, δ_{ct}, one for each county-year combination. This number of fixed effects leads to a regression with 11,977 explanatory variables. With a model of this size, classical estimation methods impose very taxing computational requirements. For instance, OLS would compute the model parameters $\delta=(D^TD)^{-1}D^TY$ by storing and manipulating in memory a covariate matrix D with 1.35×10^{11} elements (504GB of memory). The burden of the operations involved can be significantly alleviated using partitioned regression methods and parallel computing. In these methods, a master computing node coordinates K slave nodes. The master partitions the D and Y matrices into K pairs of smaller submatrices D_k and Y_k, k=1...K. Each slave node k computes the products $D_k^TD_k$ and $D_k^TY_k$ and returns them to the master node, which uses them together with special (sometimes iterative) algorithms to compute D^TD, D^TY, and $\hat{\delta}_{ct}$. We use the high-performance SAS implementation of these methods to obtain estimates of the 11,977 fixed effects δ_{ct}.

Turning to the second step, to obtain the parameters of interest, β, we implement the covariance estimator (Hsiao 2014, p. 62) by computing $\Delta\Delta\hat{\delta}_{ct} = \hat{\delta}_{ct} - \overline{\hat{\delta}_{\cdot t}} - \overline{\hat{\delta}_{c\cdot}} + \overline{\overline{\hat{\delta}}}_{\cdot\cdot}$ and then estimating

$$\Delta\Delta\hat{\delta}_{ct} = \beta_0\Delta\Delta AVGINCOME_{ct} + \beta_1\Delta\Delta GINI_{ct} + \sum_d \omega_d\Delta\Delta x_{ct,d} + \Delta\Delta\varepsilon_{ct}. \tag{7}$$

We prefer the covariance estimator to estimating the actual fixed effects for all counties and years, otherwise the number of explanatory variables would increase by 1,718 (the number of fixed effects in the second step: 7 years plus 1,711 counties). The smaller number of explanatory variables and the reduction in the number of observations to 11,977 make the estimation of (7) computationally tractable. Still, the implementation of FGLS methods robust to non-i.i.d. errors involves computing and manipulating several

matrices (e.g., $\Omega = \Delta\Delta\varepsilon_{ct}\Delta\Delta\varepsilon_{ct}^{T}$) with more than 1.1×10^{8} elements (547MB of memory) that altogether require about 30GB of memory. We thus perform the second estimation step on a high-performance computing cluster.[8]

4. Main Results

We evaluate several estimation approaches to handle the potential correlation and clustering of the model residuals in equation (7) and present the results in Table 3. We use (i) OLS, (ii) FGLS with heteroscedasticity-robust standard errors (Greene 2003, p. 314) and serially correlated errors, (iii) FGLS with only spatially correlated errors, and (iv) FGLS combining (ii) and (iii). For serial correlation, we consider an autoregressive structure of order one (AR1) (implemented as in Hsiao 2014, p. 66). For spatial correlation, we adopt an error structure of the form $\varepsilon = \rho W\varepsilon + \eta$, where W is a matrix of adjacencies (see Anselin 1988, chapter 8). For the models without spatially correlated errors (i and ii), the z-score of the standardized Moran I statistic is greater than 1.96, suggesting that spatial correlation is indeed present. In addition, accounting for spatial effects reduces the magnitude of the coefficients of the statistically significant terms.

We note that the signs, significance levels, and orders of magnitude are stable for the coefficients on average income and Gini across each of these specifications, with positive coefficients for average income and negative coefficients for Gini. Thus, we observe that product assortment increases with increasing average income and decreases with greater income dispersion. The consistency of the results across columns provides some confidence that our main findings are robust to different error specifications. In the

[8] As an alternative to the two-stage approach, we could estimate a simplified version of model (7) for each product category in one single stage and obtain category-specific estimates. Such an approach, however, is demanding on the data because it involves estimating the roughly 1.5×10^{6} distinct elements of the Ω matrix on the 11,977 observations of each category. The estimates of the Ω matrices and the standard errors may thus be inaccurate and lead to incorrect inferences. Furthermore, not all categories are present in all counties, and thus the statistical power may not be sufficient to estimate the model parameters for a number of categories. We nonetheless obtained single-stage estimates for 20 randomly selected categories within the Dry Grocery department and find that the results for *GINI* are consistent with our main results though significantly noisier. Out of the 20 coefficients of *GINI* in these 20 regressions, 12 are negative and significant, one was positive and marginally significant, and the remaining seven were not significant.

analysis that follows, we elect to use the results from Column (4), the spatial-heteroskedastic-AR1 FGLS specification because it is the most conservative approach.

We compare the size of the effect of average income on assortments with the size of the effect of the Gini on assortments by computing the elasticities of changes in UPC counts with respect to changes in average income and changes in the Gini. Evaluating at the county-level Gini average of .435 yields an elasticity of assortment size with respect to Gini of -.11,[9] while the elasticity with respect to average real income (evaluated at the county-level mean of $45,456) is .17. This indicates that the sensitivity of changes in UPC counts with respect to changes in the Gini index is of the same order of magnitude as the impact of changes in the average income level, though changes in average income are a more important driver of changes in product assortment.[10]

Robustness

We further investigate the stability of our results and whether the inclusion of the 2008-2009 economic recession in our data affected our result. To do so, we conduct separate analyses for the 2007-2010 and 2010-2013 periods. We estimate the model for 2007-2010 without the variables *NPARENTS* and *NSTORES*, because Nielsen did not consistently report store and retailer codes for each of the years during this time period. We estimate the model for 2010-2013 with and without these variables to show that including them does not affect the estimates of our variables of interest. The results appear in Table 4. The coefficients on average income and the Gini index are very close across the two time periods. Thus, the recession does not seem to be driving our results and our main effects are independent of the presence of a recession. Further, the similarity in the coefficients of average income and the Gini index across the time periods demonstrates

[9] The elasticity measured directly from the log-log specification (see Model 5 in Table W2-1) is also -0.11.

[10] These estimates are consistent with Cirera and Masset (2010), who conducted a simulation analysis to relate changes in the Gini index to changes in food demand based on Engel-curve estimates. They estimate an elasticity of food demand with respect to Gini of -.20. Our results may be lower because Cirera and Masset (2010) used data from India to guide their simulation, whereas we study the U.S. Also, we include food and non-food items; indeed, when we run the analysis only on the Dry Grocery department, we get an elasticity of -0.15, which is closer to Cirera and Masset's estimate. To the extent that income elasticities for food and non-food categories in the U.S. are higher than food in India (i.e., necessity categories play a smaller role), we may expect an attenuated relationship between changes in Gini and category demand and therefore smaller shifts in UPC and brand availability by manufacturers and retailers.

the robustness of our results and supports the notion that the effect is not driven by changes in the distribution of consumer preferences because preferences changes are less likely in shorter periods of time (Bronnenberg, Dubé, and Gentzkow 2012). One might initially assume that the similarity of the coefficients in the two time periods occurs because counties experience a common trend in UPC counts and income across the two time periods. However, this assumption does not hold. In particular, the correlation of the change in UPC counts in 2007-2010 with the change in UPC counts in 2010-2013 is close to zero (-0.06).[11] Thus, the estimation of the effects of the changes in average income and the Gini index on changes in UPC counts is driven by a different pattern of changes in the two time periods. Despite this fact, the estimates of the effects are almost the same, which gives strong confidence in the results.

We next explore how different functional specifications for the income terms affect our results. We estimate seven models (each includes all of the control variables), as reported in Table W2-1 of Web Appendix W2. We consider models where the Gini enters with a linear effect, a quadratic effect, a linear effect with an interaction of Gini with average income, a logarithmic effect of Gini, and a linear effect of Gini plus polynomial effects of average income. Overall, the similarity in the estimated coefficients of average income and Gini across columns and the results of the heteroscedasticity-robust Wald and Lagrange multiplier (LM) tests (Wooldridge 2010, p. 62) suggests that the linear model performs well and that the Gini does not proxy for higher-order terms of average income. We also compute various Palma-type measures (Palma 2011), which are defined as the ratios of the income shares of high earners over the income shares of low earners and are often used as measures of income inequality. We then test how these Palma measures affect product assortment, and obtain results that have consistent signs and significances on the effects of average income and income inequality (see Table W2-2 in Web Appendix W2).

The above analysis demonstrates a robust positive (negative) relationship between the changes in average income (income dispersion) in a county and changes in product offerings. Yet, some readers may ask whether these effects occur because consumers migrate to stores not covered in the dataset, such as

[11] The correlation in the changes in the Gini index in each time period is -0.43, reflecting a reversion toward the long-term Gini trend after the recession.

dollar stores, whose footprint increased during the period of time analyzed, or online grocery retailers. To assess this possibility, we use Nielsen's Homescan panel data to analyze household-level expenditures across stores. Households in the Homescan data set report all of their purchases across all stores regardless of whether or not the stores collaborate with Nielsen. Hence, we can use the data set to assess whether households shifted their expenditure to stores not considered in the main analysis. We compute, for each household, the proportion of spending at Nielsen stores over the years considered in our main analysis and then aggregate across households. The results, depicted in Figure 5, suggest that consumers did not migrate towards other brick and mortar stores not in the main analysis during this time period. Likewise, we note that during the period of study online grocers accounted for a very small share of the market and thus migration to online channels is unlikely to explain the results. Even after several years of fast growth, online groceries still accounted for less than 2% of the entire market in terms of revenues in 2018 (Rogers 2018).

Another question that some readers may ask is whether the county-level effects reflect reduced differentiation across stores and not the demand-driven mechanisms we propose. We thus measure the extent to which assortments shift within each store. To do so, we use demographic and purchase data in the Nielsen consumer panel to construct measures of the income inequality and demographic diversity of the patrons of individual stores. We likewise use store data to compute store-level measures of product assortment (see Web Appendix W2 for details). The results (also reported in Web Appendix W2) indicate that store assortment size increases with average income and decreases with the Gini index. The coefficients are not statistically different from the results from the main model, although point estimates are somewhat smaller than those in our main analysis. Nevertheless, the store-level effects validate the notion that the main results are driven, at least partially, by the demand mechanisms we discussed. Further, these store-level results do not rely on the census data at all, so the census reporting of demographics on a 3-year rolling window does not significantly impact our main results in Table 3.

Underlying Mechanisms

In the introduction, we presented different theories about why average income and income dispersion could affect assortment size. Here, we investigate which of the mechanisms discussed earlier may be behind the observed effects.

A key reason why one might obtain a negative coefficient on the Gini index is that income dispersion could affect category spending at supermarkets. To gauge this effect, we first model category sales (*CATEGORYSALES*) as a function of average income and Gini, while controlling for other demographics that proxy for unobserved heterogeneity. In particular, we let

(8) $$\log(CATEGORYSALES_{ict}) = \alpha_0 + \alpha_1 AVGINCOME_{ct} + \alpha_2 GINI_{ct} + X_{ct}\boldsymbol{\omega} + \tilde{\xi}_{ict},$$

where X_{ct} are the same county-level controls used in the main analysis. Estimation results for this model appear in Column (1) of Table 5. They show that increases in average income lead to increased spending and, consistent with Engel's Law for expenditure (Engel 1857), increases in Gini lead to decreased spending.

We next extend our main model (equation 4) to include the logarithm of category sales as an additional control. Category sales could theoretically be endogenous to the assortment size because increased assortment may induce additional consumption. Hence, we rely on a control function approach (Wooldridge 2010) to address the potential endogeneity bias. For instruments, we assume that the choice to eat outside the home affects the level of spending in a supermarket, but has no direct relationship to the size of the assortments offered in a market outside of this spending mechanism. Accordingly, we collect county-level data on the number and payroll of different types of prepared-food establishments and use these variables as instruments for grocery category sales (see Web Appendix W1). We use a large number of instruments because the expected correlations are relatively weak for each individual instrument. While using a large number of weak instruments may circumvent this issue, it may also induce over-parameterization (Bai and Ng 2010). Thus, we perform step-wise regressions on the original instruments, select the instruments that contribute most to the regression's R^2, and then factor analyze the selected

instruments (see Bai and Ng 2010). Lastly, we select the 14 factors[12] that explain most variation (which we call *IVFACTOR1* to *IVFACTOR14*), stack them into the vectors Z_{ct}, and add them to equation (8), giving

(9) $\log(CATEGORYSALES_{ict}) = \alpha_0 + \alpha_1 AVGINCOME_{ct} + \alpha_2 GINI_{ct} + X_{ct}\boldsymbol{\lambda} + Z_{ct}\boldsymbol{\psi} + \xi_{ict}$.

Following the control function approach, we use the estimated residuals $\hat{\xi}_{ict}$ as an additional covariate in the main model

(10) $\log(nUPCs_{ict}) = \beta_0 AVGINCOME_{ct} + \beta_1 GINI_{ct} + \tau_{it} + \gamma_{ic} + \varepsilon_{ct} + X_{ct}\boldsymbol{\omega} + \beta_2 \log(CATEGORYSALES_{ict}) + \beta_3 \hat{\xi}_{ict} + u_{ict}$,

to account for the correlation between the two equations and to address the endogeneity of category sales.

Estimates for model (9) appear in Column (2) of Table 5 (the coefficients of the instrument factors are reported in Table W3-1 of Web Appendix W3). Estimates for different specifications of model (10) appear in Table 6. Column (1) presents results of a base model without income distribution or endogeneity controls. Column (2) presents results of a similar model that includes average income and Gini. Column (3) presents estimates for a model that includes sales and the estimated residuals $\hat{\xi}_{ict}$ but not average income or Gini. Column (4) presents the full model, which includes sales, the estimated residuals, and both average income and Gini. Where included, the coefficient of the residuals $\hat{\xi}_{ict}$ is statistically significant, although the coefficients for average income and Gini shift very little.

The coefficient of average income is positive even after controlling for category spending. However, the coefficient is much smaller in Table 6 than in Table 3, reflecting that most of the effect of higher income on assortment comes from increased spending. The remaining positive effect may reflect that demand for variety increases with income, as reported by Chai, Rohde, and Silber (2015) and Carlson et al. (2015). Wald and Lagrange Multiplier tests support using the models that include average income and Gini (reported at the bottom of Table 6).

The estimated coefficient for the Gini index is negative, indicating that higher income dispersion leads retailers to offer smaller assortments even after controlling for spending. Again, the coefficients are

[12] This was the number of factors that led to the highest adjusted R^2 in the regression of sales on the instruments.

much smaller in Table 6 than in Table 3, indicating that a large part of the effects of income dispersion on assortment size occurs through category spending. One might expect income dispersion to have a positive effect on assortment size after controlling for customer spending given the theories on product differentiation outlined in the introduction (e.g., Gabszewicz and Thisse 1979; Shaked and Sutton 1983). On the other hand, as also noted in the introduction, an increase in the Gini index can reflect a "hollowing out" of the middle class. In this case, the lower income group becomes more homogenous, and a higher Gini index ironically represents a decrease in customer heterogeneity. Table 2 presents some evidence of this, where we observe that the Gini index is *negatively* correlated with the range of incomes for the bottom 80% of the population (*INCRANGE0TO80*, the level corresponding to the 80th percentile of income).[13] Thus, even when overall income inequality rises, as measured by the Gini index at the county level, the dispersion of income pertaining to the bottom four quintiles contracts.

We briefly provide some support for increases in Gini being associated with less dispersion in the incomes of the lower and middle-income patrons of the Nielsen stores in our dataset. We replicate the analysis for model (10) but replace *GINI* by *INCRANGE0TO80*.[14] The coefficient on this measure of income dispersion in Column (5) of Table 6 is both positive and significant. This indicates that increases in the income heterogeneity of the "non-rich" induce assortment growth. For robustness, we also report results for the dispersion variables *INCRANGE0TO60* and *INCOMERANGE0TO95,* also in Table 6 (other ranges are considered in Table W2-3 of Web Appendix W2). The coefficient of *INCRANGE0TO60* is very similar to that of *INCRANGE0TO80*, while the measure capturing the income range up to the 95th percentile is negatively signed, like the Gini index. Together, these results reinforce that greater income heterogeneity of the "non-rich" induces assortment growth, consistent with the predictions from the theoretical literature about product differentiation (e.g., Gabszewicz and Thisse 1979; Shaked and Sutton 1983).

---

[13] The Census Bureau only provides a few summary statistics of income, including the Gini index and the cutoff for each quintile of income.

[14] The correlation between *INCRANGE0TO80* and *AVGINCOME* is .97 and thus we repeat the analysis dividing *INCRANGE0TO80* by *AVGINCOME*. The correlation between this ratio and *AVGINCOME* is only -0.35 and the ratio is a unit-free measure of variance (just like the Gini is unit-free), analogous to the coefficient of variation. The regression coefficient of the ratio is positive and statistically significant (results available on request).

5. Managerial Implications

We find that the number of product offerings increases with average income and declines with income dispersion and related these effects to existing theories. We now examine which products are removed from shelves when average income and income dispersion change. This analysis is important for manufacturers so they may understand which products to defend—or seek to introduce—in various markets. For example, manufacturers may be able to use the insights from our findings together with local area data to improve the targeting of promotional activity. Markets where changes to income and Gini augur favorably for more assortment size could be targeted for new product introductions, while manufacturers may instead work to defend their turf in areas where assortment is expected to decline. Because the changes in income and Gini vary substantially across local areas, there is a potential opportunity to target promotional activity accordingly.

How do changes in income distribution affect product differentiation?

We begin by investigating the effect of changes in average income and income Gini on the extent of horizontal and vertical differentiation in grocery categories. We follow Jaravel (2018) in using unit price to determine the vertical quality of UPCs, and segment each product category into 12 quality tiers (we control for geographic and temporal variation in prices, as described in Web Appendix W4). We then propose two rough measures of differentiation: the number of non-empty tiers in each category, *nTIERs*, and the average number of UPCs per tier across non-empty tiers, *AVGTIERUPCs*. *nTIERs* is meant to proxy for the extent of vertical differentiation in the category, while *AVGTIERUPCs* is meant to proxy for the extent of horizontal differentiation. We then replicate the main analysis but we use log(*nTIERs*) and log(*AVGTIERUPCs*) as the dependent variables instead log(*nUPCs*). The estimated coefficients of average income, presented in Table 7, are positive and significant, indicating that greater average income increases both horizontal and vertical differentiation. The estimated coefficients of income Gini are negative and also significant, suggesting that dispersion reduces both horizontal and vertical differentiation. The effects on horizontal differentiation appear to be almost twice as large as the effects on vertical differentiation. Thus,

most of the paring down that occurs with decreasing average income and increasing income dispersion is due to reductions in horizontal differentiation.

Which products are most sensitive to changes in the income distribution?

We explore which brands win or lose when the income distribution changes. We first test whether private label or branded products lose more products. We do this by running separate analyses using only the number of private-label UPCs or only the number of branded UPCs. (All products are classified as either branded or private label.) The results are reported in Columns (3) and (4) of Table 7. For both models, we find that the coefficients for average income and Gini are very similar in magnitude and statistical significance. Thus, changes in average income and Gini affect the assortment of private labels and branded products to about the same extent. This may reflect the comparable market shares that private labels now command versus branded products and the improvements in quality that have broadened the appeal of private labels to a wide range of income levels.

We next examine whether large or small brands (brands with market shares above or below the average share in their categories) lose more UPCs when average income decreases and income dispersion increases.[15] The results appear in Columns (5) and (6) of Table 7. We observe larger coefficients on average income and Gini for large brands vs. small brands. Because the dependent variables are logs of the number of products, the coefficients can be interpreted as percentage changes in the number of UPCs. Thus, we observe that large brands lose a greater share of UPCs than small brands. An implication of this is that large brands may need to pay more attention to shifts in the income distribution when defending their turf or seeking to expand it.

6. Conclusion

The purpose of this paper has been to study the empirical relationship between changes in the distribution of incomes and changes in the available assortment of grocery products, captured at the local level. We examined 944 product categories in 1,711 U.S. counties (which accounted for 94% of the U.S. population)

[15] We use Nielsen's definition of brands for this analysis.

over seven years (from 2007 to 2013). We employed a two-step econometric approach to obtain feasible GLS estimates using more than 11 million observations. The results provide strong empirical evidence that decreases in average income and increases in income dispersion lead to decreases in assortments, holding constant demographics and other factors. We demonstrate that most of this effect occurs through the changes in overall spending that occur due to changes in the income distribution. The effects of how income dispersion affects customer spending is consistent with Engel's law for expenditure. The concavity of the Engel curve implies that, as inequality increases, consumers whose incomes rise increase their consumption but not as much as consumers whose incomes fall decrease theirs. As a result, an increase in the Gini index, holding average income constant, leads to a reduction in category demand for grocery products. This reduction, in turn, leads retailers to reduce the assortment available to consumers.

In further analyses, we examine which products are removed when income inequality increases. We find that the diminished assortment mostly reflects reduced horizontal differentiation rather than reduced vertical differentiation. We also show that larger brands lose a greater fraction of their UPCs than smaller brands. Our results give insight for manufacturers to commit resources to defend their shelf space by investing in relationships with retailers that currently stock their products in areas where incomes are shrinking or income dispersion is growing—especially for top manufacturers in a category. On the other hand, while income dispersion grew nationally during the time span of our data, the Gini index actually fell in about one-third of U.S. counties. Numerous local areas also had assortments increase, representing opportunities for manufacturers to expand their presence on retail shelves. We hope that managers may be able to harness the predictive power of income distribution changes to improve how they tailor product lines at the local level.

The increase or decrease in assortment that arises from changes in the income distribution also have implications for consumer welfare (see Israilevich 2004; Dukes et al. 2009; Jaravel 2018). In particular, consumers who lose access to their preferred items may be most adversely affected (Broniarczyk, Hoyer, and McAlister 1998). Broniarczyk et al. (1998) also show that consumer perceptions of assortment

attractiveness fall with reductions in the count of stock keeping units but that the continued presence of favorite items and the maintenance of similar category space can mitigate this effect.

Further, our analysis sheds light on the potentially unintended consequences of policy interventions. For example, the U.S. government recently increased the minimum number of offerings per category required for retailers to participate in the food stamp program (London 2017). A long-term implication of our results is that consumers' ability to use food stamps may decline in areas where falling average income and rising income dispersion leads retailers to reduce their assortments and possibly stop participating in the food stamp program.[16]

Limitations and Future Research

As an observational study, our results are valid to the extent that our identification assumptions hold. For instance, we have implicitly assumed that the effect of product assortment on residential choice is negligible. This allows us to assess the relationship between income distribution shifts and assortment by controlling for population size, number of stores, and other variables, rather than by modeling a completely endogenous system. This is the same approach used in recent work that isolates the relationships between product assortment and consumer demographics (Hwang, Bronnenberg, and Thomadsen 2010) and between CPG availability and population size (Handbury and Weinstein 2014). An inability to directly control for household relocation is a limitation of working with store-level scanner data. Future research may find alternative strategies to address this issue.

Our findings are also limited to non-perishable categories of consumer packaged goods. While these products are important to the overall economy, it would be of interest to explore the relationship between average income, Gini, and assortment size in other product classes. Indeed, taking a broader view of food and grocery consumption to include non-necessities outside the Nielsen data could potentially reverse the direction of the effects we find for Gini. We hope that the modeling approach we developed and the

[16] We make no judgment about whether the current level of assortment offerings is too high, too low, or just right. Rather, our goal is simply to demonstrate the link between income distribution changes and product assortment. While changes in assortment may have welfare implications (both positive and negative), we do not interpret any set of results in a normative way.

connections we have drawn to relevant theory (Engel's Law for expenditure, Engel's Law for variety, and product differentiation) will motivate further research into the relationships among income distribution changes and marketing outcomes.

References

Allcott, Hunt, Rebecca Diamond, and Jean-Pierre Dubé (2017), "The Geography of Poverty and Nutrition: Food Deserts and Food Choices across the United States," Working paper No. w24094. National Bureau of Economic Research.

Anselin, Luc (1988), *Spatial Econometrics: Methods and Models*, Springer Netherlands.

Athey, Susan, and Guido Imbens (2002), "Identification and Inference in Nonlinear Difference-In-Differences Models," *Econometrica*, 74(2), 431-497.

Bai, Jushan, and Serena Ng (2010), "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26(6), 1577-1606.

Barrios, Thomas, Rebecca Diamond, Guido W. Imbens, and Michal Kolesár (2012), "Clustering, Spatial Correlations, and Randomization Inference," *Journal of the American Statistical Association*, 107(498), 578-591.

Berman, Jillian (2014), "How Inequality Is Hurting Whole Foods," *Huffington Post* (Jul 7).

Broniarczyk, Susan M., Wayne D. Hoyer, and Leigh McAlister (1998), "Consumers' Perception of the Assortment Offered in a Grocery Category: The Impact of Item Reduction," *Journal of Marketing Research*, 35(2), 166-176.

Bronnenberg, Bart J., Jean-Pierre H. Dubé, and Matthew Gentzkow (2012), "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review,* 102(6), 2472-2508.

Bronnenberg, Bart J. and Vijay Mahajan (2001), "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables," *Marketing Science*, 20(3), 284-299.

Carlson, Kurt A., Jared Wolfe, Simon J. Blanchard, Joel C. Huber, and Dan Ariely (2015), "The Budget Contraction Effect: How Contracting Budgets Lead to Less Varied Choice," *Journal of Marketing Research*, 52(3), 337-348.

Chai, Andreas, Nicholas Rohde, and Jacques Silber (2015), "Measuring the Diversity of Household Spending Patterns," *Journal of Economic Surveys*, 29(3), 423-440.

Cirera, Xavier and Edoardo Masset (2010), "Income Distribution Trends and Future Food Demand," *Philosophical Transactions of the Royal Society B*, 365, 2821-2834.

Correia, Sergio (2017), "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator," Working Paper, Boston College.

Cowell, Frank A., and Emmanuel Flachaire (2013), "Statistical Methods for Distributional Analysis," in *Handbook of Income Distribution*, Vol. 2, Anthony B. Atkinson and Francois Bourguignon, eds., North-Holland Amsterdam, 359-465.

Dabla-Norris, Era, Kalpana Kochhar, Nujin Suphaphiphat, Frantisek Ricka, and Evridiki Tsounta (2015), "Causes and Consequences of Income Inequality: A Global Perspective," Research report, International Monetary Fund.

Derby, Michael (2015), "Janet Yellen: Economic Inequality Long an Interest of the Fed," *Wall Street Journal*, Apr 2.

Draganska, Michaela, Daniel Klapper, and Sofia B. Villas-Boas (2010), "A Larger Slice or a Larger Pie? An Empirical Investigation of Bargaining Power in the Distribution Channel," *Marketing Science*, 29(1), 57-74.

Du, Rex Y., and Wagner A. Kamakura (2008), "Where Did All that Money Go? Understanding How Consumers Allocate Their Consumption Budget," *Journal of Marketing*, 72(6), 109-131.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi (2018), "Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession," *Marketing Science*, 37(1), 22-53.

Dukes, Anthony J., Tansev Geylani, and Kannan Srinivasan (2009), "Strategic Assortment Reduction by a Dominant Retailer," *Marketing Science*, 28(2), 309-319.

Engel, Ernst (1857), "Die Produktions-und Konsumptionsverhältnisse des Königreichs Sachsen," *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Ministeriums des Innern*, 8, 1-54.

Gabszewicz, Jaskold J. and Jacques-François Thisse (1979), "Price Competition, Quality, and Income Disparities," *Journal of Economic Theory,* 20, 340-359.

Greene, William H. (2003), *Econometric Analysis,* 5th ed., Prentice Hall Upper Saddle River, NJ.

Grewal, Dhruv, Michael Levy, Anuj Mehrotra, and Arun Sharma (1999), "Planning Merchandising Decisions to Account for Regional and Product Assortment Differences," *Journal of Retailing*, 75(3), 405-424.

Handbury, Jessie and Weinstein, David E. (2014), "Goods Prices and Availability in Cities," *Review of Economic Studies*, 82(1), 258-296.

Hansen, Christian B. (2007), "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects," *Journal of Econometrics*, 140(2), 670-694.

Hsiao, Cheng (2014), *Analysis of Panel Data*. Cambridge, UK: Cambridge University Press.

Hwang, Minha, Bart J. Bronnenberg, and Raphael Thomadsen (2010), "An Empirical Analysis of Assortment Similarities Across US Supermarkets," *Marketing Science*, 29(5), 858-879.

Israilevich, Guillermo (2004), "Assessing Supermarket Product-Line Decisions: The Impact of Slotting Fees," *Quantitative Marketing and Economics*, 2(2) 141-167.

Jaravel, Xavier (2018), "The Unequal Gains from Product Innovation: Evidence from the US Retail Sector," *Quarterly Journal of Economics*, 134(2), 715-783.

Kadiyali, Vrinda, Pradeep Chintagunta, and Naufel Vilcassim (2000), "Manufacturer-Retailer Channel Interactions and Implications for Channel Power: An Empirical Investigation of Pricing in a Local Market," *Marketing Science*, 19(2), 127-148.

Kamakura, Wagner A. (2014), "What Happened to the American Middle Class? Class and Consumption in America," Working paper, Rice University.

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

Kunreuther, Howard (1973), "Why the Poor May Pay More for Food: Theoretical and Empirical Evidence," *Journal of Business*, 46(3), 368-383.

Lancaster, Kelvin (1990), "The Economics of Product Variety: A Survey," *Marketing Science*, 9(3), 189-206.

Leigh, Andrew (2007), "How Closely Do Top Income Shares Track Other Measures of Inequality?," *The Economic Journal*, 117(524), F619-F633.

Lewbel, Arthur (2008), "Engel Curve," *The New Palgrave Dictionary of Economics*, 2nd edition, edited by Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

Lodd, Jit (2016), "Effective Premiumization Strategies in CPG can Achieve 10–15 Percent Topline Growth," Research report, Information Resources Inc.

London, Lori (2017), "Meeting Stocking Requirements to Maintain SNAP Eligibility,", (accessed Sept. 14, 2017), [available at https://www.goebt.com/snap-eligibility-stocking-requirements/].

Orhun, Yesim and Mike Palazzolo (2019), "Frugality is Hard to Afford," *Journal of Marketing Research*, 56(1), 1-17

Palma, José Gabriel (2011), "Homogeneous Middles vs. Heterogeneous Tails, and the End of the 'Inverted-U': It's all about the Share of the Rich," *Development and Change*, 42(1), 87-153.

Ries, Al (2013), "Why JCP, Walmart and Others Fail at Changing Their Spots," *Advertising Age*, May 14.

Ren, Charlotte R., Ye Hu, Yu Hu, and Jerry Hausman (2011), "Managing Product Variety and Collocation in a Competitive Environment: An Empirical Investigation of Consumer Electronics Retailing," *Management Science*, 57(6), 1009-1024.

Rogers, David (2018), "Is Online Grocery Really Going to Explode in the US?" Supermarket News, Aug 24, (accessed May 2, 2019), [available at https://www.supermarketnews.com/online-retail/viewpoint-online-grocery-really-going-explode-us].

Rosenblum, Paula and Steve Rowen (2008), "Customer-Centric Merchandising: Driving Differentiation through Localization," Benchmark report, Retail Systems Research.

Salazar, David (2014), "Relevance, Urgency Prompt Purchasing," *Drug Store News*, August 25, 144-146.

Shaked, Avner and John Sutton (1983), "Natural Oligopolies," *Econometrica*, 51(5), 1469-1484.

Silva, Santos J.M.C. and Silvana Tenreyro (2006), "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641-658.

Simanis, Erik and Duncan Duke (2014), "Profits at the Bottom of the Pyramid," *Harvard Business Review*, 92(10), 86-93

Talukdar, Debabrata (2008), "Cost of Being Poor: Retail Price and Consumer Price Search Differences Across Inner-City and Suburban Neighborhoods," *Journal of Consumer Research*, 35(3), 457-471.

Van Herpen, Erica and Rik Pieters (2002), "The Variety of an Assortment: An Extension to the Attribute-Based Approach," *Marketing Science,* 21(3), 331-341

Watson, Elaine (2014), "Sales Growth in US Yogurt Category Will Likely 'Venture into Negative Territory in 2015'," (accessed August 10, 2016), [available at www.foodnavigator-usa.com/content/view/print/1006624].

Wooldridge, Jeffrey M. (2010), *Econometric Analysis of Cross Section and Panel Data,* MIT Press Cambridge.

Table 1: DATA DESCRIPTION

	Mean	St. Dev.	Min	Median	Max
log(*nUPCs*)	3.01	0.78	-0.17	3.12	4.72
AVGINCOME ($100000)	0.45	0.11	0.23	0.43	1.05
GINI	0.44	0.03	0.33	0.44	0.60
INCRANGE0TO80 ($100000)	0.88	0.20	0.41	0.84	2.00
INCRANGE0TO60 ($100000)	0.57	0.14	0.25	0.55	1.41
INCRANGE0TO95 ($100000)	1.48	0.35	0.77	1.39	4.43
log(*CATEGORYSALES*)	8.90	1.52	4.39	8.80	14.12
UNEMPLOYMENT (10)	0.01	0.003	0.00	0.01	0.03
HOMEVALUE ($1000000)	0.12	0.07	0.02	0.10	0.72
HIGHSCHOOL (10)	0.03	0.01	0.01	0.03	0.06
POPULATION (1000000)	0.17	0.41	0.02	0.06	9.95
HISPANIC	0.05	0.11	0.00	0.00	1.00
EDUCDVRSTY	0.12	0.04	0.03	0.12	0.32
AGEDVRSTY	0.06	0.005	0.06	0.06	0.13
HHSIZEDVRSTY	0.12	0.02	0.05	0.12	0.34
ETHNICDVRSTY	0.77	0.17	0.29	0.81	1.00
FOREIGN	0.01	0.05	0.00	0.00	0.48
NPARENTS	4.79	2.50	1.00	4.00	14.00
NSTORES	20.16	42.10	1.00	7.00	705.00

Note: Detailed variable descriptions appear in Web Appendix W1

Table 2: CORRELATIONS

	log(*nUPCs*)	*AVGINCOME*	*GINI*	*INCRANGE0TO80*	*INCRANGE0TO60*	*INCRANGE0TO95*	log(*CATEGORYSALES*)	*UNEMPLOYMENT*	*HOMEVALUE*	*HIGHSCHOOL*	*POPULATION*	*HISPANIC*	*EDUCDVRSTY*	*AGEDVRSTY*	*HHSIZEDVRSTY*	*ETHNICDVRSTY*	*FOREIGN*	*NPARENTS*	*NSTORES*
log(*nUPCs*)	1																		
AVGINCOME	0.35	1																	
GINI	0.14	-0.11	1																
INCRANGE0TO80	0.43	0.97	-0.11	1															
INCRANGE0TO60	0.39	0.96	-0.28	0.95	1														
INCRANGE0TO95	0.45	0.93	0.12	0.93	0.87	1													
log(*CATEGORYSALES*)	0.35	0.39	0.17	0.39	0.34	0.45	1												
UNEMPLOYMENT	0.15	-0.39	0.24	-0.30	-0.36	-0.26	-0.01	1											
HOMEVALUE	0.33	0.79	0.00	0.77	0.73	0.76	0.42	-0.21	1										
HIGHSCHOOL	-0.49	-0.44	-0.31	-0.50	-0.42	-0.57	-0.31	-0.08	-0.44	1									
POPULATION	0.31	0.33	0.19	0.33	0.27	0.38	0.96	0.02	0.38	-0.28	1								
HISPANIC	0.22	0.20	0.13	0.22	0.17	0.25	0.31	0.08	0.27	-0.38	0.36	1							
EDUCDVRSTY	-0.38	-0.11	-0.31	-0.18	-0.11	-0.26	-0.16	-0.27	-0.18	0.88	-0.17	-0.36	1						
AGEDVRSTY	0.07	0.16	0.11	0.17	0.15	0.18	0.05	-0.10	0.16	-0.34	0.05	0.13	-0.20	1					
HHSIZEDVRSTY	0.16	-0.05	0.46	-0.03	-0.12	0.06	0.23	0.12	0.04	-0.19	0.24	0.14	-0.14	0.11	1				
ETHNICDVRSTY	-0.28	-0.20	-0.31	-0.23	-0.15	-0.28	-0.32	-0.17	-0.22	0.41	-0.34	-0.37	0.39	-0.18	-0.33	1			
FOREIGN	0.25	0.44	0.18	0.45	0.38	0.48	0.65	-0.02	0.55	-0.28	0.68	0.34	-0.12	0.06	0.26	-0.36	1		
NPARENTS	0.71	0.49	0.09	0.49	0.45	0.50	0.47	-0.06	0.47	-0.34	0.43	0.26	-0.18	0.07	0.19	-0.34	0.41	1	
NSTORES	0.37	0.35	0.21	0.34	0.29	0.41	0.94	0.01	0.36	-0.29	0.94	0.33	-0.17	0.04	0.28	-0.37	0.64	0.53	1

Table 3: MAIN RESULT AND SELECTION OF RESIDUAL SPECIFICATION

	Dependent variable: log(*nUPCs*)			
	(1)	(2)	(3)	(4)
	OLS	FGLS	FGLS	FGLS
		het, AR(1)	SP	het, SP, AR(1)
UNEMPLOYMENT	-0.929	-0.257	0.136	0.460
	(0.687)	(0.738)	(0.668)	(0.711)
HOMEVALUE	0.136	0.215***	0.093	0.181**
	(0.096)	(0.072)	(0.106)	(0.083)
HIGHSCHOOL	9.351***	6.990***	8.281***	6.374***
	(1.359)	(1.458)	(1.309)	(1.367)
POPULATION	-0.778***	-0.745***	-0.880***	-0.858***
	(0.219)	(0.126)	(0.215)	(0.151)
HISPANIC	-0.002	-0.007	0.005	-0.008
	(0.032)	(0.040)	(0.030)	(0.036)
EDUCDVRSTY	-1.249***	-0.939***	-1.206***	-0.898***
	(0.179)	(0.191)	(0.174)	(0.186)
AGEDVRSTY	-0.706	0.107	-0.174	0.588
	(0.965)	(0.699)	(0.907)	(0.635)
HHSIZEDVRSTY	-0.076	-0.020	-0.104	-0.027
	(0.145)	(0.175)	(0.135)	(0.152)
ETHNICDVRSTY	0.102***	0.056*	0.094***	0.057**
	(0.029)	(0.031)	(0.027)	(0.028)
FOREIGN	0.115	0.106***	0.109	0.099***
	(0.087)	(0.028)	(0.082)	(0.027)
NPARENTS	0.340***	0.342***	0.336***	0.330***
	(0.022)	(0.030)	(0.020)	(0.028)
NSTORES	0.220***	0.186***	0.196***	0.170***
	(0.020)	(0.024)	(0.018)	(0.022)
AVGINCOME	6.325***	5.195***	4.508***	3.762***
	(0.676)	(0.700)	(0.650)	(0.649)
GINI	-0.354***	-0.357***	-0.260***	-0.257***
	(0.073)	(0.083)	(0.069)	(0.077)
Fixed effects	Included	Included	Included	Included
Observations	11,977	11,977	11,977	11,977
Moran I Z	10.896	13.532	1.771	1.771
Within Adjusted R^2	0.103	0.091	0.102	0.092

Notes: *p<0.10; **p<0.05; ***p<0.01; het - heteroscedasticity-consistent standard errors; AR(1) - estimates corrected for serial correlation of order 1; SP - estimates corrected for spatial correlation.

Table 4: ROBUSTNESS TO ECONOMIC CONDITIONS

	Dependent variable: log(*nUPCs*)		
	2007-2010 period	2010-2013 period	2010-2013 period
	(1)	(2)	(3)
UNEMPLOYMENT	0.506	2.996***	2.789***
	(1.016)	(0.931)	(0.837)
HOMEVALUE	0.254**	0.208	-0.054
	(0.105)	(0.227)	(0.197)
HIGHSCHOOL	9.833***	-1.355	-1.120
	(1.864)	(2.532)	(2.161)
POPULATION	-1.010***	-1.012***	-0.588**
	(0.179)	(0.260)	(0.297)
HISPANIC	-0.061	0.044	0.018
	(0.057)	(0.047)	(0.038)
EDUCDVRSTY	-1.239***	-0.078	-0.010
	(0.242)	(0.403)	(0.344)
AGEDVRSTY	0.496	3.543**	3.053*
	(0.709)	(1.769)	(1.663)
HHSIZEDVRSTY	-0.011	-0.108	-0.049
	(0.227)	(0.199)	(0.178)
ETHNICDVRSTY	-0.012	0.142***	0.079*
	(0.035)	(0.050)	(0.044)
FOREIGN	0.104***	0.022	0.087**
	(0.028)	(0.037)	(0.042)
NPARENTS			0.300***
			(0.029)
NSTORES			0.166***
			(0.018)
AVGINCOME	2.610***	3.315***	3.338***
	(0.891)	(0.933)	(0.831)
GINI	-0.294**	-0.226**	-0.243***
	(0.118)	(0.102)	(0.093)
Fixed effects	Included	Included	Included
Observations	6,848	6,844	6,844
Within Adjusted R^2	0.010	0.011	0.177

Notes: *p<0.10; **p<0.05; ***p<0.01; *NSTORES* and *NPARENTS* do not exhibit sufficient variation for identification over years 2007-2010 due to Nielsen data collection practices prior to 2010. Results for the years 2010-2013 are consistent with and without these two variables.

Table 5: EFFECT OF CHANGES IN INCOME DISTRIBUTION ON CATEGORY SALES

	Dependent variable: log(*CATEGORYSALES*)	
	(1)	(2)
Constant	8.165***	7.446***
	(0.191)	(0.191)
UNEMPLOYMENT	14.177***	16.987***
	(2.232)	(2.239)
HOMEVALUE	1.277***	1.884***
	(0.161)	(0.173)
HIGHSCHOOL	-73.225***	-51.822***
	(2.820)	(2.910)
POPULATION	-0.014	-0.630***
	(0.048)	(0.105)
HISPANIC	0.500***	0.432***
	(0.063)	(0.063)
EDUCDVRSTY	8.722***	6.097***
	(0.465)	(0.468)
AGEDVRSTY	-7.629***	-4.080***
	(1.373)	(1.361)
HHSIZEDVRSTY	5.201***	5.937***
	(0.450)	(0.458)
ETHNICDVRSTY	-0.452***	-0.480***
	(0.044)	(0.043)
FOREIGN	-2.707***	-1.569***
	(0.192)	(0.196)
NPARENTS	0.363***	0.355***
	(0.003)	(0.003)
NSTORES	0.008***	0.007***
	(0.0005)	(0.001)
AVGINCOME	4.896***	7.766***
	(1.164)	(1.185)
GINI	-0.691***	-0.641***
	(0.213)	(0.209)
Instrument Factors	Not Included	Included
Observations	11,977	11,977
Within Adjusted R^2	0.814	0.826

Notes: *p<0.10; **p<0.05; ***p<0.01. Estimated coefficients of instrument factors appear in Table W3-1 of Web Appendix W3.

Table 6: EFFECT OF CHANGES IN CATEGORY SALES ON ASSORTMENT

	Dependent variable: log(*nUPCs*)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
UNEMPLOYMENT	0.876*	1.263**	1.103**	1.568***	1.479***	1.479***	1.378***
	(0.503)	(0.507)	(0.507)	(0.511)	(0.513)	(0.514)	(0.509)
HOMEVALUE	-0.194***	-0.266***	-0.162***	-0.232***	-0.207***	-0.210***	-0.210***
	(0.040)	(0.043)	(0.042)	(0.044)	(0.044)	(0.044)	(0.044)
HIGHSCHOOL	2.193**	2.391***	1.364	1.399	1.750*	1.535*	1.422
	(0.873)	(0.875)	(0.909)	(0.911)	(0.921)	(0.911)	(0.905)
POPULATION	-0.378***	-0.376***	-0.355***	-0.348***	-0.360***	-0.355***	-0.344***
	(0.047)	(0.046)	(0.047)	(0.046)	(0.046)	(0.047)	(0.046)
HISPANIC	0.00004	-0.002	0.007	0.006	0.008	0.007	0.007
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
EDUCDVRSTY	-0.328***	-0.341***	-0.229*	-0.223*	-0.272**	-0.245*	-0.230*
	(0.121)	(0.121)	(0.125)	(0.125)	(0.126)	(0.125)	(0.124)
AGEDVRSTY	1.526***	1.623***	1.446***	1.531***	1.492***	1.507***	1.458***
	(0.470)	(0.473)	(0.470)	(0.474)	(0.469)	(0.470)	(0.470)
HHSIZEDVRSTY	-0.076	0.006	0.009	0.116	0.077	0.083	0.064
	(0.091)	(0.092)	(0.096)	(0.097)	(0.098)	(0.097)	(0.097)
ETHNICDVRSTY	-0.014	-0.014	-0.021	-0.023*	-0.023*	-0.022*	-0.022
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
FOREIGN	0.038*	0.041*	0.017	0.015	0.016	0.015	0.016
	(0.022)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
NPARENTS	0.039***	0.040***	0.045***	0.046***	0.046***	0.046***	0.045***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
NSTORES	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
log(*CATEGORYSALES*)	0.439***	0.438***	0.424***	0.420***	0.421***	0.420***	0.422***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$\hat{\xi}_{ict}$			0.015***	0.019***	0.017***	0.018***	0.017***
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
AVGINCOME		0.187***		0.198***	0.041	0.083	0.182***
		(0.045)		(0.045)	(0.054)	(0.055)	(0.048)
GINI		-0.129**		-0.139**			
		(0.054)		(0.054)			
INCRANGE0TO80					0.083***		
					(0.030)		
INCRANGE0TO60						0.074*	
						(0.042)	
INCRANGE0TO95							-0.013*
							(0.007)
Fixed effects	Included	Included	Included	Included	Included	Included	Included
Observations	11,977	11,977	11,977	11,977	11,977	11,977	11,977
Within Adjusted R^2	0.594	0.594	0.594	0.594	0.594	0.594	0.594
Wald test p-val		0.0001[a]	0.002[a]	0.000[b]	0.0001[b]	0.001[b]	0.001[b]
LM test p-val		0.000[a]	0.026[a]	0.000[b]	0.000[b]	0.000[b]	0.0001[b]

Notes: *p<0.10; **p<0.05; ***p<0.01. [a] relative to Column (1); [b] relative to Column (3)

Table 7: EFFECTS OF INCOME DISTRIBUTION ON CATEGORY STRUCTURE

	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
	log(*nTIERs*)	log(*AVG TIERUPCs*)	log(*nUPCs*) private	log(*nUPCs*) national	log(*nUPCs*) small brands	log(*nUPCs*) large brands
UNEMPLOYMENT	0.147	0.206	0.240	0.445	0.377	0.420
	(0.204)	(0.430)	(0.581)	(0.687)	(0.448)	(0.804)
HOMEVALUE	0.022	0.073*	-0.009	0.181**	0.170***	0.206**
	(0.017)	(0.039)	(0.073)	(0.082)	(0.054)	(0.094)
HIGHSCHOOL	1.945***	3.007***	5.776***	5.426***	3.887***	7.607***
	(0.397)	(0.868)	(1.154)	(1.322)	(0.852)	(1.553)
POPULATION	-0.156***	-0.174***	-1.878***	-0.482***	-0.606***	-0.991***
	(0.017)	(0.035)	(0.134)	(0.147)	(0.098)	(0.171)
HISPANIC	-0.001	-0.008	-0.007	-0.008	-0.003	-0.012
	(0.009)	(0.019)	(0.029)	(0.035)	(0.023)	(0.041)
EDUCDVRSTY	-0.314***	-0.450***	-0.740***	-0.798***	-0.564***	-1.065***
	(0.055)	(0.120)	(0.156)	(0.180)	(0.115)	(0.211)
AGEDVRSTY	0.008	0.305	0.435	0.512	0.347	0.667
	(0.166)	(0.363)	(0.545)	(0.628)	(0.414)	(0.719)
HHSIZEDVRSTY	-0.012	-0.078	0.157	-0.078	-0.041	-0.012
	(0.041)	(0.085)	(0.121)	(0.148)	(0.095)	(0.172)
ETHNICDVRSTY	0.008*	0.017*	0.018	0.054**	0.039**	0.065**
	(0.005)	(0.010)	(0.022)	(0.027)	(0.018)	(0.032)
FOREIGN	0.049***	0.051***	0.088***	0.098***	0.081***	0.101***
	(0.008)	(0.017)	(0.026)	(0.027)	(0.019)	(0.030)
NPARENTS	0.017***	0.038***	0.252***	0.335***	0.214***	0.368***
	(0.001)	(0.002)	(0.021)	(0.028)	(0.018)	(0.032)
NSTORES	-0.0001**	-0.0005***	0.134***	0.155***	0.101***	0.194***
	(0.0001)	(0.0001)	(0.017)	(0.022)	(0.014)	(0.025)
AVGINCOME	0.119***	0.209***	3.414***	3.434***	2.458***	4.317***
	(0.019)	(0.041)	(0.537)	(0.630)	(0.407)	(0.735)
GINI	-0.087***	-0.161***	-0.257***	-0.228***	-0.171***	-0.296***
	(0.022)	(0.047)	(0.061)	(0.075)	(0.048)	(0.087)
Fixed effects	Included	Included	Included	Included	Included	Included
Observations	11,977	11,977	11,977	11,977	11,977	11,977
Within Adjusted R^2	0.050	0.047	0.088	0.091	0.094	0.091

Notes: *p<0.10; **p<0.05; ***p<0.01.

Figure 1: HISTORICAL LOWER INCOME LIMITS FOR INCOME GROUPS IN THE U.S.

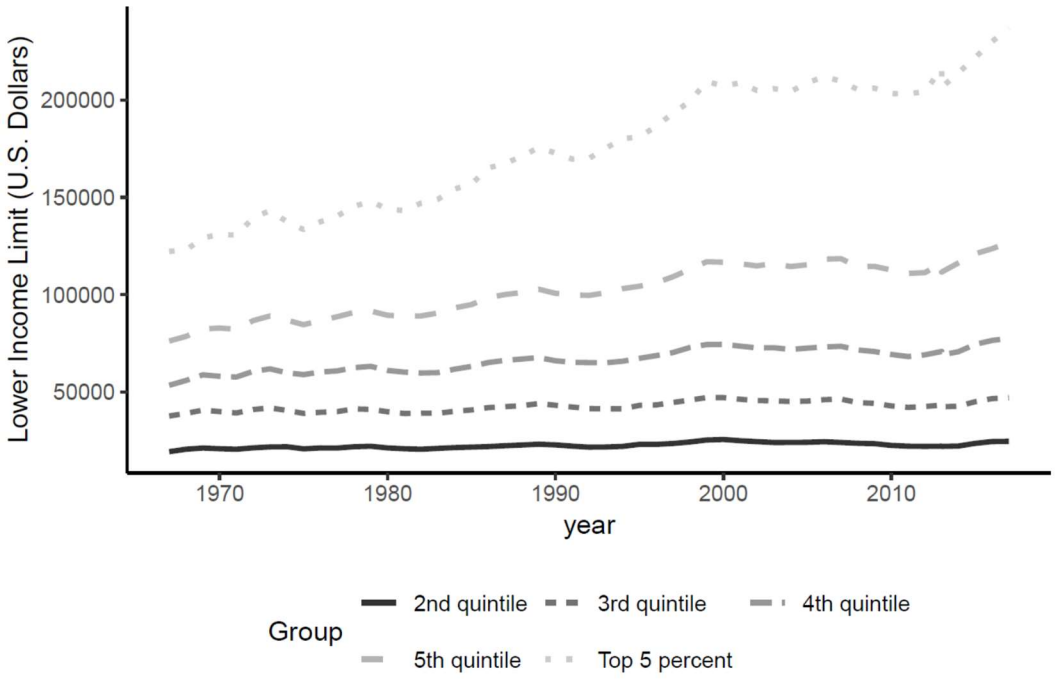
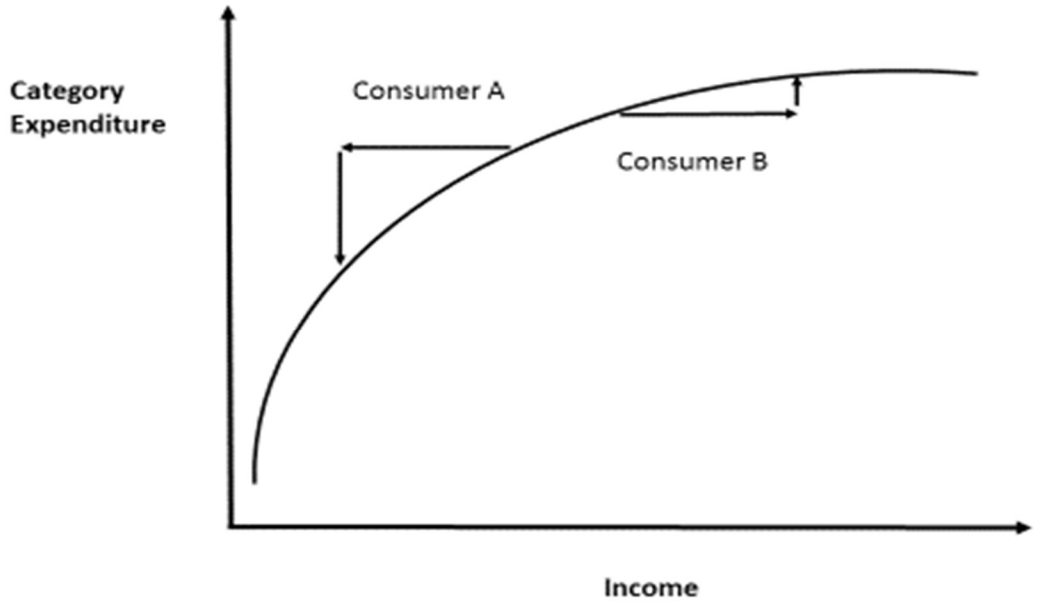


Figure 2: INCOME DISTRIBUTION CHANGES AND THE ENGEL CURVE

A



B

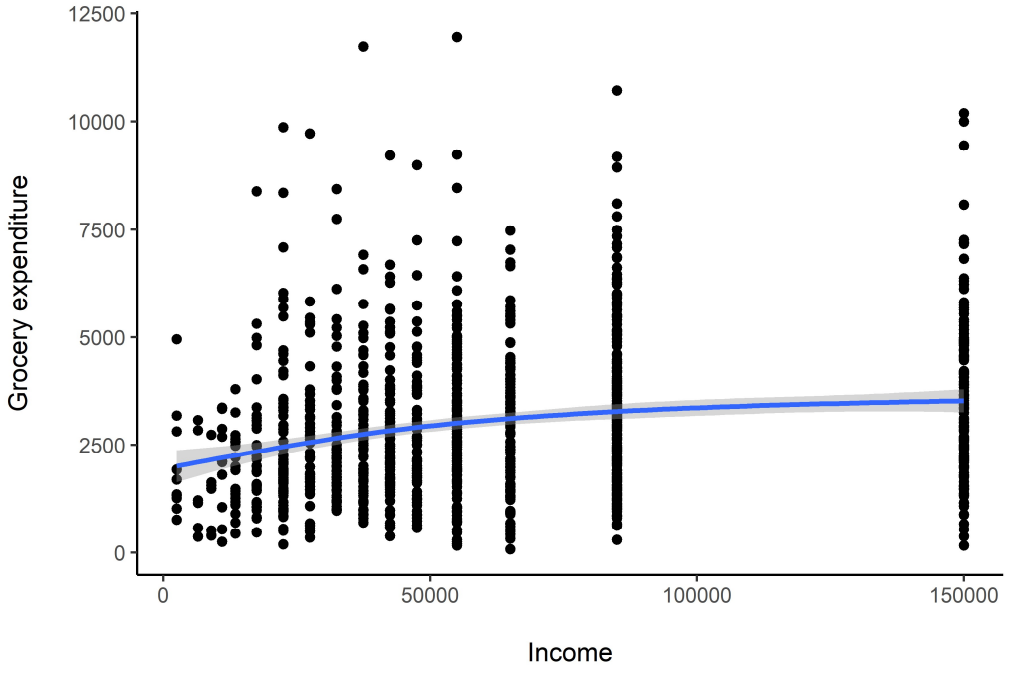


Figure 3: SAMPLE VARIATION OF MAIN VARIABLES

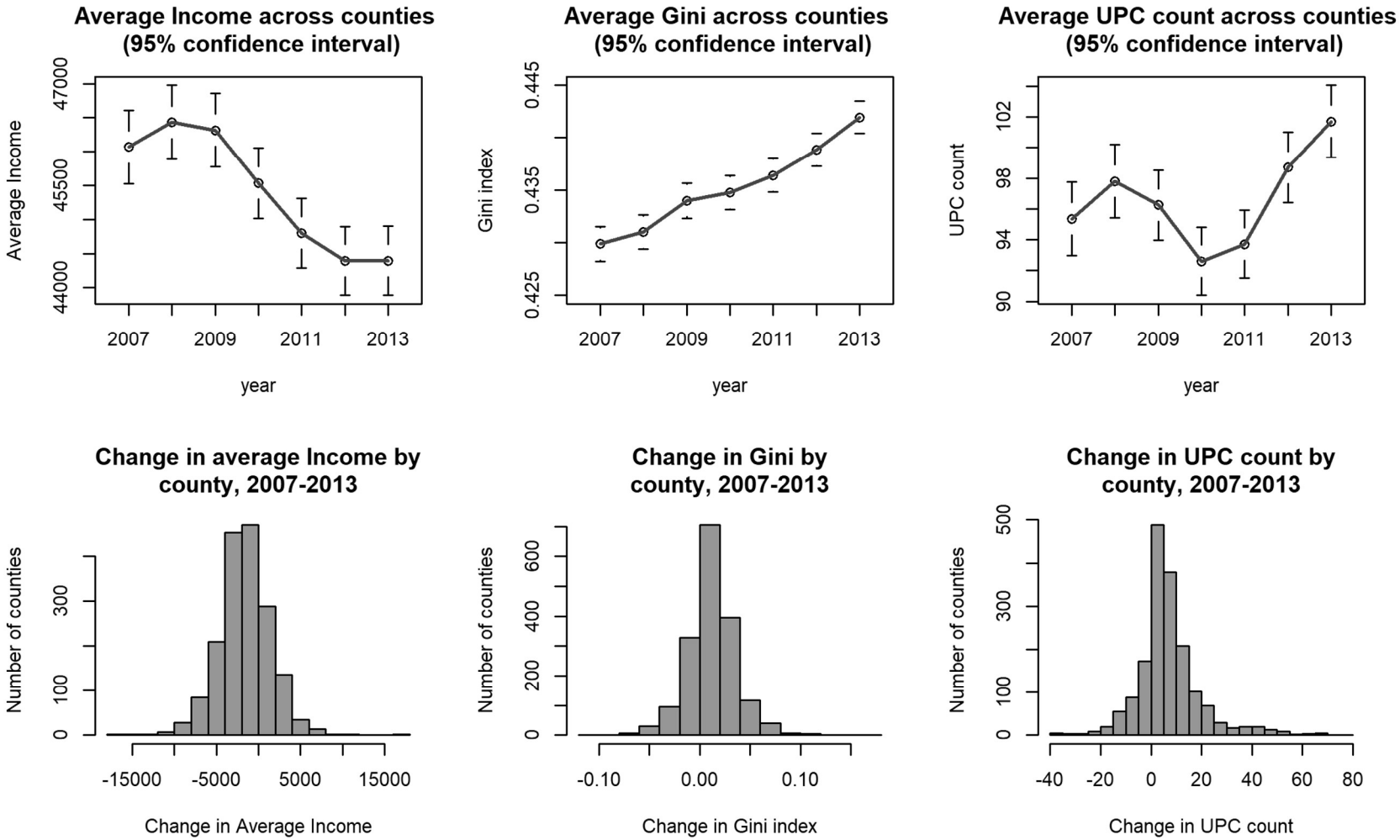


Figure 4: CHANGES IN PRODUCT (UPC) COUNTS AND UPC COUNTS PER 10,000 DOLLARS IN INCOME FOR COUNTIES WITH DECREASED AND INCREASED AVERAGE INCOME AND INCOME GINI, 2007 TO 2013

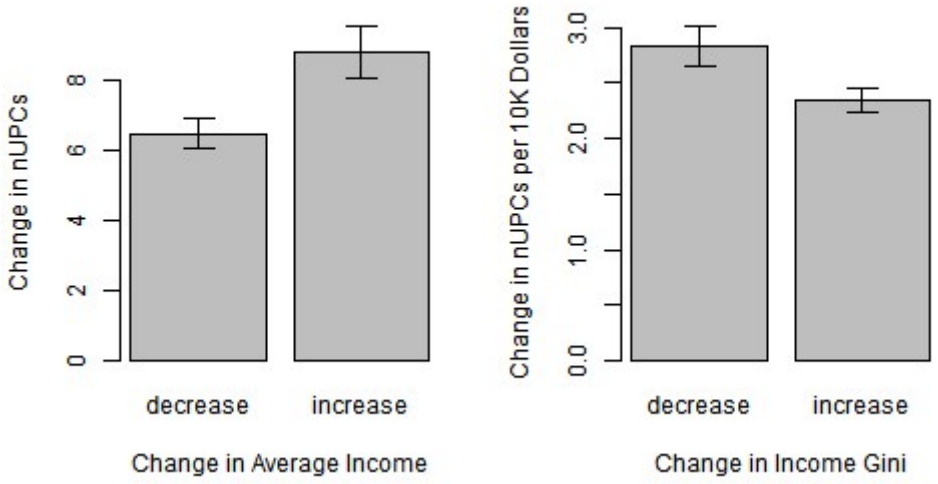


Figure 5: PROPORTION OF HOUSEHOLD EXPENDITURE IN NIELSEN STORES