The Unequal Gains from Product Innovations: Evidence from the US Retail Sector*

Xavier Jaravel, Stanford University

June 26, 2016

Note to NBER Summer Institute participants:
a revised version of the paper will be uploaded on July 10

Abstract

Using detailed product-level data in the retail sector in the United States from 2004 to 2013, this paper shows that product innovations disproportionately benefited high-income households due to increasing inequality and the endogenous response of supply to market size. Annualized quality-adjusted inflation was 0.65 percentage points lower for high-income households, relative to low-income households. Using national and local changes in market size driven by demographic trends plausibly exogenous to supply factors, the paper provides causal evidence that a shock to the relative demand for goods (1) affects the direction of product innovations, and (2) leads to a decrease in the relative price of the good for which demand became relatively larger (i.e. the long-term supply curve is downward sloping). A calibration shows that this channel can explain most of the observed difference in quality-adjusted inflation rates across the income distribution.

JEL codes: E31, I31, I32, O30, O31, O33

*I am particularly indebted to Daron Acemoglu, Philippe Aghion, Raj Chetty, Ed Glaeser, Larry Katz, Pete Klenow, Josh Lerner, Marc Melitz and Andrei Shleifer for extensive advice and comments. I also thank Ufuk Akcigit, David Autor, Matt Basilico, Shai Bernstein, Michael Best, Laura Blattner, Nick Bloom, Kirill Borusyak, Michael Boskin, Diego Comin, Gary Chamberlain, Varanya Chaubey, Gabriel Chodorow-Reich, John Coglianese, Tilman Dette, Mark Duggan, Max Eber, Josh Feng, Peter Ganong, Andy Garin, Matt Gentzkow, Robert Hall, Ben Hebert, Nathan Hendren, Leander Heldring, Erik Hurst, Simon Jaeger, Chad Jones, Max Kasy, Bill Kerr, David Koretz, David Laibson, Danial Lashkari, Jing Li, Hannes Mamberg, Sendhil Mullainathan, Emi Nakamura, Thomas Piketty, Benjamin Pugsley, Dan Pollman, Alexandra Roulet, Jann Spiess, Stefanie Stancheva, Jeremy Stein, Larry Summers, Colin Tomkins-Bergh, Chris Tonetti, Gabriel Unger, Heidi Williams, Alwyn Young and seminar participants at Harvard, Stanford, the College de France, the Kauffman Foundation and the Econometric Society North American Summer Meeting for thoughtful discussions and comments. I gratefully acknowledge financial support from the the HBS doctoral office, the Kauffman Foundation, the Stanford Institute for Economic Policy Research and the Washington Center for Equitable Growth.

1 Introduction

The rising level of nominal income inequality in the United States over the past thirty years has been a key topic of interest for academics and policymakers. The effect of skill-biased technical change in this process has been widely studied: innovations such as the diffusion of information and communications technologies have affected the relative price of skills and resulted in higher nominal income inequality. Much less attention has been paid to how price changes in the product market and the introduction of new products may differentially affect households at different points of the income distribution. Yet it is well-known that preferences are non-homothetic. Depending on their income level, households consume very different goods and services. Due to price changes in the product market over time, as well as changes in product variety, trends in real inequality may therefore differ from trends in nominal inequality. Product innovations may play a central role in this process by increasing the variety and quality of goods available in specific consumer segments, as well as by driving down the price of existing products in these segments due to increased competitive pressure. This paper shows the relevance of this hypothesis in the US retail sector over the past ten years, a large sector accounting for over 25% of the US economy.

I investigate this question in two steps. First, in the "measurement" part of the paper, I show that in the retail sector over the past ten years the quality-adjusted price index of high-income households rose substantially *slower* than that of low-income households, which amplified inequality. I build income-group-specific quality-adjusted price indices using detailed product-level data, in which I observe consumption patterns across income groups, price changes for all products available in consecutive years (inflation) and changes in product variety (product entry and exit). Second, in the "mechanism" part of the paper, I find that firms' equilibrium response to changes in demand across the income distribution explains why the price index of high-income consumers rose slower than that of the low-income. Specifically, my analysis shows that because demand from the high-income grew faster during this period, firms strategically introduced more new products catering to these consumers, which in turn drove down the price of existing products in these segments due to competitive dynamics.[1] The retail sector is ideal to conduct this investigation because it accounts for a large share of US GDP, rich data is available, and the notion of product (barcode) is well defined.

[1] A particularly good example illustrating this idea is the market for snacks. In recent years, meat snacks have grown tremendously, especially premium beef jerky - with sustained double-digit growth for over five years nationwide. Premium beef jerky is a high-protein, low-fat and low-calorie snack - a practical and healthy snack that particularly appeals to young and high-income households. The branding of premium beef jerky is fundamentally different from that of traditional jerky - favorite of truckers and staple of gas-station checkouts - and so is its production process. In particular, many of the varieties of premium beef jerky are fully organic - for instance, beef jerky made from 100% grass-fed cattle from networks of small family farms. The so-called "jerky renaissance" is largely driven by demand. It is answering the demand of high-income consumers concerned with healthy living and eager to support a sustainable, more humane agriculture. And it is taking place in a broader context of increased demand for snacks - a Nielsen survey found that one in ten Americans say they eat snacks instead of meals - and for proteins - according to the NPD group, more than half of Americans say they want more protein in their diet. The competition for the premium beef jerky market has intensified in recent years, with an ever increasing number of small, local players but also with the entry of established companies through acquisitions. For instance, Krave, one of the early players in premium jerky who led the market in the late 2000s, was acquired in 2015 by Hershey's, the largest chocolate manufacturer in North America. Accordingly, premium beef jerky prices have fallen and varieties have increased. Similar - although less spectacular - dynamics are visible in other segments of the snack industry, like hummus and protein bars, but not so in segments catering to lower-income consumers, like chips, bars and nuts.

In the first part of the paper, I establish two new facts about inflation and increasing product variety across the household income distribution in the US retail sector from 2004 to 2013. I find that higher-income households experienced lower inflation and a faster increase in product variety than more modest households. The magnitude of these effects is large: over this period, the average annual inflation rate was 0.65 percentage point lower for households making more than $100,000 a year, compared with households making less than $30,000.[2] These results are very robust and hold before, during and after the Great Recession across product groups for a wide variety of price indices. They are based on detailed product-level data from the Nielsen Homescan Consumer Panel and Retail Scanner datasets, which are representative of the retail sector as a whole (which itself represents 40% of household expenditures on goods and 16% of household total expenditures). Whether similar results hold in other sectors of the economy is uncertain, but the analysis delivers a general methodological lesson for the measurement of inflation by statistical agencies: I show that the difference in inflation rates across the income distribution can be accurately measured only with product-level data. Indeed, a large share of the inflation difference between income groups occurs within detailed product categories, which cannot be captured by price series based on data aggregated at a level similar to what the Bureau of Labor Statistics and other statistical agencies currently use. These findings challenge the result from the existing literature that inflation is similar across the income distribution[3] and suggest that trends in real inequality may be diverging from trends in nominal inequality. Collecting product-level data is key to accurately measure this divergence. This has important potential policy implications given the indexation of many government transfers.

In the second part of the paper, I examine whether the equilibrium response of supply to faster growth of demand from high-income consumers can explain the new facts on differential inflation and increase in product variety across the income distribution. It is a natural hypothesis to investigate because it is well known that in recent decades the share of national income accruing to high-income consumers has steadily increased (e.g. Song et al. 2016). To do so, I introduce a micro-founded model featuring monopolistic competition with variable elasticity of substitution preferences that differ across income groups, which generates a set of precise predictions that I take to the data and for which I find strong support. Intuitively, firms respond to changes in relative market size by skewing product introductions toward market segments that are growing faster. This process leads to a decrease in the price of existing products in the fast-growing market segments because increased competitive pressure from new products pushes markups down, and also because firms endogenously engage in process innovations lowering the marginal cost of these products. In my data, product groups catering to higher-income households grow faster and have a higher rate of product introduction, as well as lower inflation on existing products, which provides suggestive evidence in support of the theory but does not establish causality from demand to supply.

To test the causal claim that increases in demand lead to a fall in inflation and an increase in product

[2] As discussed in Section 4, increasing product variety is valuable on its own, but empirically most of the welfare difference between households across the income distribution are captured by price changes in the basket of products that are available across years.

[3] See Section 4 for a detailed discussion of how my results relate to the literature.

variety, I rely on two complementary identification strategies. First, I use shifts in the national age and income distribution between 2004 and 2013 to estimate the causal effect of changes in the number of consumers (market thickness) in a given part of the product space on inflation and product innovations. This research design is similar in spirit to Acemoglu and Linn (2004). Second, I introduce a novel research design exploiting variation in food stamp policy across US states between 2000 and 2007 to trace out the impact of changes in per capita spending on inflation and product innovation. Both research designs provide variation in demand plausibly exogenous to supply factors.[4] They are complementary because the first is based on variation in the number of consumers and the second on variation in per capita spending, which in principle could have different effects on the equilibrium.[5]

Taken together, my results show that in response to growing demand the equilibrium price of existing products falls and the rate of introduction of new products increases, and that these effects are sufficiently strong to explain the divergence in price indices across the income distribution. According to my point estimates, a 1 percentage point increase in demand leads to a 10 basis point decline in inflation and a 35 basis point increase in spending on new products. In line with the model, the magnitude of the effect is similar regardless of whether the change in demand comes from a change in the number of consumers or in per capita spending. In simple calibrations, I show that these effects are large enough to explain the new facts documented in the first part of the paper. In other words, these results suggest that absent the endogenous response of supply to market size effects, there would not have been a substantial difference in inflation nor in the rate of increase in product variety across the income distribution. This analysis has important implications for the endogenous growth literature, by providing evidence for endogenous product innovations across detailed product categories. It is also relevant for the trade literature and the debate on the role of markups in the gains from trade, because I establish empirically that the gains from increased market size are largely due to a fall in markups (consistent with the model and variable elasticity of substitution preferences). More broadly, these results are relevant for policy, given that the effectiveness of any government transfer crucially depends on the equilibrium response of supply to market size.

Overall, this paper provides new evidence challenging the existing literature primarily in two respects. First, the literature suggests that households across the income distribution tend to experience similar inflation rates (e.g. McGranahan and Paulson (2005)), except during peculiar periods like the Great Recession (Argente and Lee (2015)). Second, theoretical work has focused on the "product cycle", the idea that innovation is driven by economies of scale and allows for a trickle-down process bringing to the mass market the new products that were initially enjoyed by a select few at the top of the income distribution. In other words, innovation dynamics should contribute to lower quality-adjusted inflation for lower-income households. My findings suggest that market size effects may be a more important force, contributing to lower

[4] This version of the draft only reports results from the first research design. The research design based on changes in food stamp policy are available from the author upon request.

[5] My model relies on translog preferences with flexible preference parameters across income groups, which yields the prediction that changes in demand coming from changes in the number of consumers or from changes in per capita spending should lead to the same endogenous supply response. This prediction is borne out in the data, which rejects a broad class of monopolistic competition models using other preferences, for instance as in Zhelobodko et al. (2012).

quality-adjusted inflation for higher-income households because market size grows faster at the top of the income distribution. More generally, this paper contributes to various strands of literature studying income inequality, price indices, technical change and monopolistic competition dynamics. [6]

The remainder of this paper is organized as follows: Section 2 describes the data, Section 3 presents the first main contribution of the paper, the measurement of quality-adjusted inflation and increasing product variety across the income distribution, and Section 4 makes the second main contribution of the paper, establishing that the endogneous response of supply to market size effects is very strong. A number of theoretical results, estimation details and robustness checks are reported in appendices.

2 Data Sources

2.1 Scanner Data

The analysis is primarily based on the Nielsen Homescan Consumer Panel and Nielsen Retail Scanner datasets, which have been widely used in the literature (Einav, Leibtag and Nevo, 2008). With this data, I can track consumption from 2004 to 2013 at the product level in department stores, grocery stores, drug stores, convenience stores and other similar retail outlets across the US. The data are representative of about 40% of household expenditures on goods and 16% of total household expenditures. Appendix B presents a detailed description of the data sources.

Three features of the data are particulalry useful for my analysis. First, product-level data is available on both prices and quantities. Quantity data is rare at the product level (for instance, the Bureau of Labor Statistics (BLS) does not collect such data) but it is crucial for quality adjustment in price indices. Intuitively, observing shifts in quantities allows to directly measure substitution patterns (and thus address substitution bias, which is a core concern of the CPI produced by BLS) and to infer the quality of products given their price, their market share, and the demand system. The quantity and price data is also used to structurally estimate the relevant parameters of the demand system in Section 3. Second, the Homescan Consumer panel has information on household characteristics such as income, age, education, size, occupation, marital status and zip code. It is therefore possible to directly map products to consumer characteristics. Third, the dataset offers a good measure of product innovations, defined as the introduction of new barcodes. Broda and Weinstein (2010) provide a detailed explanation regarding why it is reasonable to assume that all goods with different UPCs differ in some substantial way that might cause consumers to pay a different price for them and that it is rare for a meaningful quality change to occur that does not result in a change of UPC.[7] In

[6] More precisely, this paper relates to at least seven strands of literature, which respectively examine nominal income inequality (Autor, Katz and Kruger (1998), Autor, Katz and Kearney (2008), Piketty (2013), Song, Price, Guvenen, Bloom and Till von Wachter (2015), Atkinson (2015)), homothetic price indices (Sato (1976), Vartia (1976), Feenstra (1992), Pakes (2003), Broda and Weinstein (2006, 2010), Erickson and Pakes (2011), Comin, Lashkari and Mestieri (2015)), non-homothetic price indices (McGranahan and Paulson (2005), Broda and Romalis (2009), Moretti (2013), Diamond (2015), Handbury (2015), Faber and Fally (2015), Argente and Lee (2015)), innovation in labor markets (Acemoglu (1996, 2002, 2007), Acemoglu and Autor (2011), Autor (2013), Autor and Dorn (2013), Bell, Chetty, Jaravel, Petkova and Van Reenen (2015)), market size effects and endogenous technical change (Acemoglu and Linn (2004)), innovation and inequality in product markets (Schumpeter (1942), Vernon (1966), Matsuyama (2002)), and trade models of monopolistic competition with free entry (Melitz (2003), Melitz and Ottaviano (2008), Zhelobodko, Kokovin, Parenting and Thisse (2012)).

[7] Broda and Weinstein (2010) make this point as follows: "Although it is difficult to enforce how a company uses a bar code,

other words, it is safe to assume that if the bar code changes, it is likely that some noticeable characteristic of the product has changed.[8] Similarly, it is possible to track products (barcodes) that are discontinued. Appendix Table 18 shows that creation is larger than destruction, i.e. new products tend to steal market shares from existing products. As will become apparent in Section 3, this displacement is indicative of biases in conventional price indexes that ignore the effects of changing quality.

Nielsen provides a detailed product hierarchy, based on where products are sold in stores. In my sample, about 3 million products (identified by their barcode, or "UPC") are classified into 10 broad departments (dry grocery, general merchandise, health and beauty care, alcoholic beverages, deli, ...), 125 more detailed product groups (grooming aids, soup, beer, pet care, kitchen gadgets, ...) and 1,075 very detailed product modules (ricotta cheese, pet litter liners, bathroom scale, tomato puree, women's hair coloring, ...). When ranking product modules by mean consumer income, the top five product modules are scotch, natural cheese, gin, fondue sauce and cookware, while the bottom five are tobacco, canned meat, taco filing, insecticide and frozen fruit drinks.

Finally, the data can be disaggregated at the level of 76 local markets, described in Appendix B. According to Nielsen, the dataset is still representative within each of the 76 markets. The data cannot reliably be disaggregated further (e.g. at the county or zip code level).

2.2 Markup Data

I also have access to weekly product-level data between January 2004 and June 2007 in 19 U.S. states, for 250 grocery stores operated by a single retail chain. This dataset contains information for 125,048 unique products (UPCs), mostly in the food and beverages categories, housekeeping supplies, books and magazines, and personal care products. Most of the stores are located in the western and eastern corridors, in the Chicago area, Colorado and Texas. For every store in every week, data is available on the price, the wholesale cost and the marginal cost of each product. I infer the markups of the retailer based on the information on the price and wholesale cost. Note that I do not measure other costs like labor, rent and utilities. In the analysis carried out in Section 4, store-year fixed effects are used to absorb these costs. The dataset also reports "adjusted gross profits" per unit for each product, defined as the net price minus the sum of wholesale costs and transportation costs plus net rebates from the manufacturer - I use this adjustment in robustness checks.

most industry experts strongly caution firms not to use the same bar code on more than one product. Doing so could cause confusion among retailers who would have trouble knowing what they were selling and for consumers whose receipts would not match their actual purchases. Similarly, firms typically do not use multiple UPCs for the same product because that makes it very difficult for retailers to reorder out of stock items. As a result, manufacturers tend to use other bar code systems for internal use and reserve the UPC for tracking products that are identical to the consumer. For example, changing the slogan on a Heinz ketchup bottle does not require a new bar code, but changing the size of the bottle does."

[8] Note that these measures of product turnover include any change in products, including those driven by changes in the size of products, their flavor, or other characteristics that can be secondary for the consumer. Nielsen provides identifiers that allows tracking barcodes that are new just because of a change in size or flavor: all of the results presented in the paper are similar when excluding these products from the definition of "new" products.

2.3 Manufacturer Data

Finally, I have obtained data from GS1, the company in charge of allocating bar codes in the US, on the universe of barcodes and manufacturers. I match the bar codes observed in the Nielsen data to manufacturers using the first few digits of the bar code - the match rate is close to 95%. Since the cutoff size for a manufacturer to appear in this dataset is to make a sale rather than an arbitrary number of workers, I can observe the full distribution of manufacturers in each product group. There are about 500 manufacturers on average in each product group, with 90 percent of the product groups having more than 200 manufacturers. The median number of products supplied by a manufacturers is 5 and the average is 14.

Consistent with the findings reported by Hottman et al. (2016), while on average half of all output in a product group is produced by just five manufacturers, around 98 percent of manufacturers have market shares below 2 percent. Thus, the typical product group is characterized by a few large manufacturers and a competitive fringe of manufacturers with very low market shares. A second important feature of the data is that even the largest manufacturers are not close to being monopolists: the largest manufacturers in a product group on average has a market share of 22 percent. The model presented in Section 4 is consistent with these patterns.

3 Measuring Quality-Adjusted Inflation Across the Income Distribution

In this section, I compute quality-adjusted inflation rates across the income distribution, taking into account the welfare gains from increasing product variety. I start with a brief reminder on non-homothetic preferences and price indices. Second, I follow standard results in the literature to derive the exact price index in my preferred demand system. Third, I present results and robustness checks for inflation on existing goods across the income distribution. I show the relevance of these results for statistical agencies like BLS and discuss differences with the existing literature. Fourth, I document the difference in changes in product variety (due to both product creation and destruction) across the income distribution. Finally, I bring together the findings on inflation on overlapping products and on product creation and destruction to compute the full quality-adjusted inflation rate.

3.1 Nonhomothetic Preferences, Product Variety and Real Inequality

The nonhomothetic nature of preferences means that the baskets of goods and services consumed by households across the income distribution systematically differ. Given that households have a taste for variety, the mapping between nominal income and utility depends on both the quality-adjusted price of products and the number of available varieties. This paper studies how the mapping between nominal income and inequality changes over time. Figure 15 illustrates this idea. In this example, the "new" mapping is an upward shift of the "old" mapping (for instance because of productivity gains), but the shift is asymmetric and benefits

higher-income households relatively more. The shift takes into account changes in the quality-adjusted price of products as well as changes in the variety of available products for each nominal income level.

This paper characterizes shifts in the mapping from nominal income to utility at various points of the income distribution using a money metric, the compensating variation. The compensating variation gives the amount of nominal income that one would need to take away from the consumer at the "new" equilibrium to make them indifferent between this new equilibrium (with the new mapping) and the "old" one (with the initial mapping). This approach provides a characterization of changes in real inequality. Given the demand system, it is possible to infer the quality of products based on their price and equilibrium market share, and to measure the gains from increasing product variety based on the share of spending on new products. The rest of this section discusses the procedure in detail and shows that the results are robust across price indices, indicating that structural assumptions about the demand system do not drive the results.

I use the term "inflation" to describe my findings throughout the paper because it is an intuitive notion, but my results are invariant to the unit of account. I document changes in the relative prices of goods that cater to high- and low-income households. These relative price changes would be unaffected by shifts in the overall level of inflation, therefore nominal indeterminacy plays no role in my findings.[9]

3.2 Overview of Methodology and Review of Basic Price Indices

The goal is to compute the cost of achieving a certain level of utility in one year relative the previous year. Such price indices are known as "exact price indices." This requires taking into account changes in product quality, product variety, as well as the optimizing behavior of consumers who may substitute from one good to another. By definition, this exercise requires taking a stance on a utility function. The role of the utility function is twofold: quantifying the impact on utility of price changes for the goods that exist across periods, but also translating into a welfare metric the patterns of product creation and destruction. In order to understand what parts of the result are driven by structural assumptions on the utility function, it is useful to split this analysis into two parts, first considering price changes on products that exist across periods and second considering changes in product variety.

First, I consider inflation on the set of products available in two consecutive years. The quality of a given product is assumed to be constant over time[10] and data is available on market shares of each product, therefore it is straightforward to compute a price index reflecting product quality and consumers' substitution behavior. Intuitively, I observe the price change for each product and I only need to decide how to weight the various products. The exact price index offers a principled way of doing so. The structural assumption on

[9] It is also useful to note that given that the set of goods is not fixed, the difference in the rates of quality-adjusted inflation experienced between high- and low-income households could be permanent. If the set of goods were fixed, the divergence in inflation rates between goods should be bounded and eventually converge to 0, otherwise in the long run all consumers, regardless of their income level, would switch to the goods with slower price increases. But since there is entry and exit, quality-adjusted inflation may be permanently lower for one income group relative to another (e.g. at any point in time the price of the products catering to the high income may remain higher than that of the products catering to the low income, but in a quality-adjusted sense the price of the high-end products may be very low).

[10] This assumption is standard in the literature: see for instance Feenstra (1994) and Broda and Weinstein (2010). It appears to be reasonable even though advertising or the introduction of complementary goods may violate it. It can be tested, and I present the results of this test in the robustness test section below.

the utility function play a minor role for the final result, as can be seen by computing standard price indices that do not have an interpretation in terms of utility but can serve as bounds by allowing for an extreme form of substitution (like the Paasche price index, which offers a lower bound on inflation) or making any substitution impossible (like the Laspeyres price index, which offers an upper bound on inflation). In addition to the exact price index derived in the following subsection, I consider the following price indices :

$$
\begin{aligned}
Laspeyres\ Index: \quad P_L &\equiv \frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} = \sum_{i=1}^{n} \frac{p_i^t}{p_i^0} s_i^0 \\
Paasche\ Index: \quad P_P &\equiv \frac{\sum_{i=1}^{n} p_i^t q_i^t}{\sum_{i=1}^{n} p_i^0 q_i^t} = \left(\sum_{i=1}^{n} \left(\frac{p_i^t}{p_i^0} \right)^{-1} s_i^t \right)^{-1} \\
Marshall-Edgeworth\ Index: \quad P_{ME} &\equiv \frac{\sum_{i=1}^{n} p_i^t (q_i^t + q_i^0)}{\sum_{i=1}^{n} p_i^0 (q_i^t + q_i^0)} \\
Walsh\ Index: \quad P_W &\equiv \frac{\sum_{i=1}^{n} p_i^t \sqrt{q_i^t q_i^0}}{\sum_{i=1}^{n} p_i^0 \sqrt{q_i^t q_i^0}} \\
Fisher\ Index: \quad P_F &\equiv \sqrt{P_L P_P} \\
Geometric\ Laspeyres\ Index: \quad P_L^G &\equiv \Pi_{i=1}^{n} \left(\frac{p_i^t}{p_i^0} \right)^{s_i^0} = exp\left(\sum_{i=1}^{n} s_i^0 \cdot log\left(\frac{p_i^t}{p_i^0} \right) \right) \\
Geometric\ Paasche\ Index: \quad P_P^G &\equiv \Pi_{i=1}^{n} \left(\frac{p_i^t}{p_i^0} \right)^{s_i^t} = exp\left(\sum_{i=1}^{n} s_i^t \cdot log\left(\frac{p_i^t}{p_i^0} \right) \right) \\
Tornqvist\ Index: \quad P_T &\equiv \Pi_{i=1}^{n} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^t + s_i^0}{2}} = exp\left(\sum_{i=1}^{n} \frac{s_i^0 + s_i^t}{2} \cdot log\left(\frac{p_i^t}{p_i^0} \right) \right)
\end{aligned}
$$

Another relevant consideration is whether or not to chain the price index. In a chain index, each link consists of an index in which each period is compared with the preceding one, the weight and price reference being moved forward each period. A chain index is therefore path dependent: it depends on the prices and quantities in all the intervening periods between the first and last period in the index series. When there is a gradual economic transition from the first to the last period, chaining is advantageous because it smoothes trends in relative prices and quantities and tends to reduce the index number spread between the various price indices listed above. But if there are fluctuations in the prices and quantities in the intervening periods, chaining may not only increase the index number spread but also distort the measure of the overall change between the first and last periods.[11] Accordingly, I present robustness check with and without chaining the indices.

Second, I follow standard techniques in the literature to provide an adjustment to the price index depending on the rate of increase in product variety. For new product and discontinued products, price relatives are not available. Intuitively, given that consumers have a taste for variety, an increase in the range of available product should lead to a decrease in the price index. Translating the increase in product variety into welfare

[11] For example, suppose all the prices in the last period return to their initial levels in period 0, which implies that they must have flucutated in between. A chain Laspeyres index will not return to 100: it will tend to be greater than 100. If the cycle is repeated with all the prices periodically returning to their original levels, a chain Laspeyres index will tend to drift further and further above 100 even though there may be no long-term upward trend in the prices. Chaining is therefore not advised when the price fluctuates.

gains requires structural assumptions. The key assumption is that within a product module, varieties are horizontally differentiated. I use two standard frameworks: nested CES utility (see next subsection) and nested translog utility (in progress). Importantly, these structural assumptions do not matter quantitatively because the elasticity of substitution of products within a module is very high, such that the gains from increasing product variety are already reflected in the prices of existing products (this is shown formally in the next subsection). The key point is that the patterns of product creation and destruction matter in general equilibrium, but their welfare effect is almost entirely taken into account in the price changes of products existing across periods. I also provide bounds showing that the patterns of product creation and destruction in the data will lower the price index more for high-income households than low-income households and, even under some violations of the structural assumptions.

Although the main text focuses on results and robustness checks across income groups (various price indices, various subsamples of products and years, etc.), I have also conducted this analysis across age groups. The results of this analysis are reported in Appendix. If find that inflation tends to be higher for older individuals, which is in line with the market size hypothesis examined in Section 4.

3.3 Estimation Framework for Nested CES Exact Price Index

The estimation framework builds on Feenstra (1994) and Broda and Weinstein (2006, 2010). I split the analysis using three representative agents, one for households making less than $30,000 a year, one for households making between $30,000 and $100,000 a year, and one for households making above $100,000. Preference parameters in my estimation framework are a flexible function of the income level, which allows for nonhomotheticities.

The remainder of this subsection shows how to derive and estimate the price index for any representative agent. I assume a nested CES utility function, following Feenstra . Product groups are indexed by g and G is the set of all product groups. The elasticity of susbtitution across product groups is ρ. The elasticity of susbtitution between product groups is $\sigma = \rho/(\rho - 1)$. The upper level utility function is:

$$U = (\sum_{g \in G} (C_{gt})^{\rho})^{\frac{1}{\rho}}$$

Composite consumption within a product group is given by:

$$C_{gt} = (\sum_{m \in M_g} (c_{mgt})^{\rho_g})^{\frac{1}{\rho_g}}$$

$\sigma_g = \rho_g/(\rho_g - 1)$ is the elasticity of substitution between product modules within product group g.

$$c_{mgt} = (\sum_{u \in U_m} (d_{umgt} c_{umgt})^{\rho_m})^{\frac{1}{\rho_m}}$$

where c_{ubgt} is the consumption of UPC u in product module m and product group g in period t. $\sigma_m = \rho_m/(\rho_m - 1)$ between UPCs within product module m. d_{umgt} is unobserved and reflects the quality of

the UPC. So we want to estimate σ and two high-dimensional sets of elasticities of substitution, $\{\sigma_g\}_g$ and $\{\sigma_m\}_m$. We expect $\sigma_m > \sigma_g$ since there is more substitution across UPCs within a module than across modules within a group.

The minumum unit cost function of the subutility function at the product module level is:

$$P_{mgt} = \left(\sum_{u \in U_{mgt}} \left(\frac{p_{umgt}}{d_{umgt}} \right)^{\sigma_m} \right)^{\frac{1}{\sigma_m}}$$

The minimum cost function at the product group level is:

$$P_{gt} = \left(\sum_{m \in M_g} (P_{mgt})^{\sigma_g} \right)^{\frac{1}{\sigma_g}}$$

And the overall price index is given by:

$$P_t = \left[\sum_g P_{gt}^{\sigma} \right]^{\frac{1}{\sigma}}$$

Consumer optimization also yields:

$$s_{umgt} = \left(\frac{p_{umgt}/d_{umgt}}{P_{mgt}} \right)^{1-\sigma_m}$$

i.e. the quality adjusted price can be backed out as follows:

$$ln \frac{p_{umgt}}{d_{umgt}} = \frac{ln(s_{umgt})}{1-\sigma_m} + ln(P_{mgt})$$

The key insight for estimation is that the share of consumption of UPC u depends directly on the quality-adjusted price. We can write the price index only in terms of prices and market shares even when goods are constantly being replaced.

If we make the assumption that product quality is constant over time ($d_{umgt} = d_{umgt-1}$) and if ignore the introduction of new products, given our assumption of a (nested) CES utility function and the results in Sato (1976) and Vartia (1976), the exact price index is:

$$P_{mg}(p_{mgt}, p_{mgt-1}, x_{mgt}, x_{mgt-1}, I_{mg}) = \Pi_{u \in I_{mg}} \left(\frac{p_{umgt}}{p_{umgt-1}} \right)^{w_{umgt}} \tag{1}$$

where $I_{mg} = I_{mgt} \bigcap I_{mgt-1}$ is the set of varieties consumed in both periods t and $t-1$. x_{mgt} and x_{mgt-1} are the cost-minimizing quantity vectors of products within module m in each of the two periods. A remarkable feature is that the price index does not depend on the unknown quality parameters d_{umgt}. We only need to compute the geometric mean of the individual variety price changes, where the weights are ideal log-change weights. These weights are computed using cost shares in the two periods and are always bounded between the shares of spending in the t and $t-1$ (in other words the price index is bounded between the geometric Paasche and Laspeyres indices described in the previous subsection):

$$s_{umgt} = \frac{p_{umgt} x_{umgt}}{\sum_{u \in I_{mg}} p_{umgt} x_{umgt}}$$

$$w_{umgt} = \frac{(s_{umgt} - s_{umgt-1})/(ln(s_{umgt}) - ln(s_{umgt-1}))}{\sum_{c \in I_{mg}} (s_{umgt} - s_{umgt-1})/(ln(s_{umgt}) - ln(s_{umgt-1}))}$$

As shown in Broda and Weinstein (2010), with change in varieties across periods the exact price index (quality-adjusted inflation) for product module m within product group g is then given by:

$$\pi_{mg}(p_{mgt}, p_{mgt-1}, x_{mgt}, x_{mgt-1}, I_{mg}) = P_{mg}(p_{mgt}, p_{mgt-1}, x_{mgt}, x_{mgt-1}, I_{mg}) \cdot \left(\frac{\lambda_{mgt}}{\lambda_{mgt-1}}\right)^{\frac{1}{\sigma_m - 1}} \quad (2)$$

with

$$\lambda_{mgt} = \frac{\sum_{u \in I_{mg}} p_{umgt} x_{umgt}}{\sum_{u \in I_{mgt}} p_{umgt} x_{umgt}} \; ; \; \lambda_{umgt-1} = \frac{\sum_{u \in I_{mg}} p_{umgt-1} x_{umgt-1}}{\sum_{u \in I_{mgt-1}} p_{umgt-1} x_{umgt-1}}$$

This result states that the exact price index with variety change is equal to the "conventional" price index multiplied by an additional term, which captures the role of new and disappearing varieties. The higher the expenditure share of new varieties, the lower is λ_{mgt} and the smaller is the exact price index relative to the conventional price index. An intuitive way to rewrite this ratio is as follows:

$$\frac{\lambda_{mgt}}{\lambda_{mgt-1}} = \frac{1 + Growth\ Rate\ of\ Spending\ on\ Overlapping\ Products_{gmt}}{1 + Growth\ Rate\ of\ Total\ Spending_{gmt}}$$

which clearly shows that a net increase in product variety (weighted by spending) drives the price index down. The price index also depends on the module-specific elasticity of substitution between varieties σ_m. As σ_m grows, the additional term converges to one and the bias goes to zero: intuitively, when existing varieties are close substitutes to new or disappearing varieties, price changes in the set of existing products already take into account the entry of more varieties.[12]

In principle, we could use the result above to compute price indices adjusted for increasing product variety over any time horizon. However, two factors make some time horizons more sensible than others in practice. First, it makes sense to define periods in years to prevent seasonal factors from driving product turnover. Thus, UPCs will be considered destroyed only if they were not purchased at any time during a yearlong period. Second, we need to decide how many years should separate the two periods. While this choice is inherently arbitrary, I decided to present results based on one-year intervals, considering other intervals in robustness checks. As mentioned earlier, a key assumption is that the taste or quality parameters for common goods must remain constant in start and end years of the sample. In fact, it may vary over short horizons due to anything that might affect demand (e.g., marketing or fashion considerations). The reason for why immutable preferences over long time horizons must be assumed when deriving price indexes is that if the utility function is changing over time for either exogenous reasons (e.g., fashion) or endogenous reasons (e.g.,

[12] One can better understand the implications of the choice of time horizon by considering an examples of how the proposition captures the impact of different types of creation and destruction, quoting from Broda and Weinstein (2010):

"Let's consider the case of a new type of sunscreen that replaces an earlier type. If the new sunscreen is just a repackaging of last year's sunscreen without a noticeably different quality or price, then, ceteris paribus, the new sunscreen will have a market share equal to that of the old sunscreen. If this is true, then the share of common goods will be unchanged and our measured quality bias from the replacement of the old model would be zero. If, instead, the new sunscreen is priced identically but is of a higher quality than the old model, then, ceteris paribus, its market share will rise. This result comes directly from the optimizing behavior of the consumer, because the new sunscreen will have a lower price per unit quality than the old sunscreen. If this is the case, the higher share of the new good relative to the old good implies that there is a "quality bias" in the conventional price index that only considers products existing across periods."

marketing) then one cannot make sensible statements about how price changes affect welfare, nor can one derive exact price indexes because identical price vectors will yield different utility levels at different times. The choice of the time horizon also matters for the magnitude of the adjustment term for increasing product variety.

Thus, we need data on quantity and price for new products, discontinued products, and products existing across periods, which is readily available in the Nielsen data. We also need to estimate the two high-dimensional sets of elasticities of substitution, $\{\sigma_g\}_g$ and $\{\sigma_m\}_m$. The main challenge for estimation of is that we want to obtain a demand and supply equation using only information on prices and quantities. The insight of Feenstra (1994) as extended by Broda and Weinstein (2006) is that although we cannot identify supply and demand, the data does tell us something about the joint distribution of supply and demand parameters. Appendix C gives details about how to derive the estimation equations.[13]

3.4 Inflation Across the Income Distribution For Goods Available in Consecutive Years

3.4.1 Results

Figure 1: Inflation Across the Income Distribution (Overlapping Products)

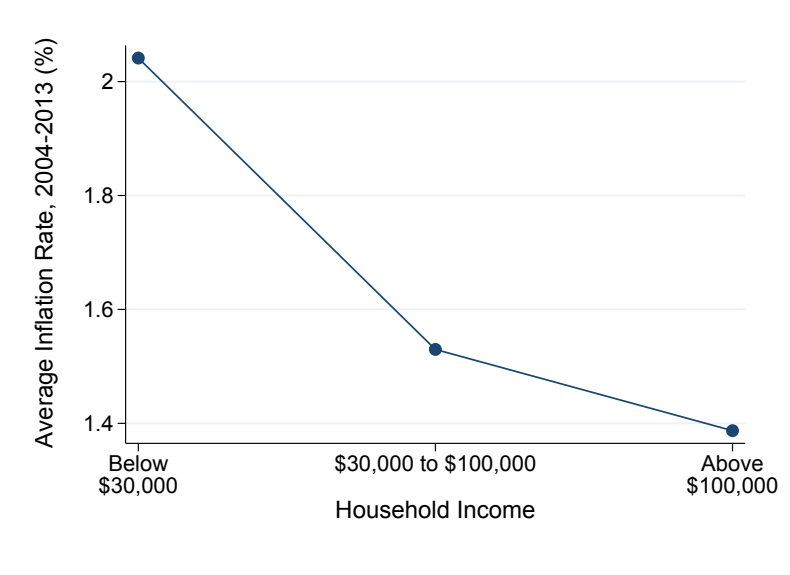


Figure 1 shows the average inflation between 2004 and 2013 on the set of overlapping products (defined as products that are available in consecutive years) for households across the income distribution. Inflation is computed using the exact price index for the nested CES utility function described in the previous subsection (without the adjustment for new and disappearing products, which is examined later in this section and does

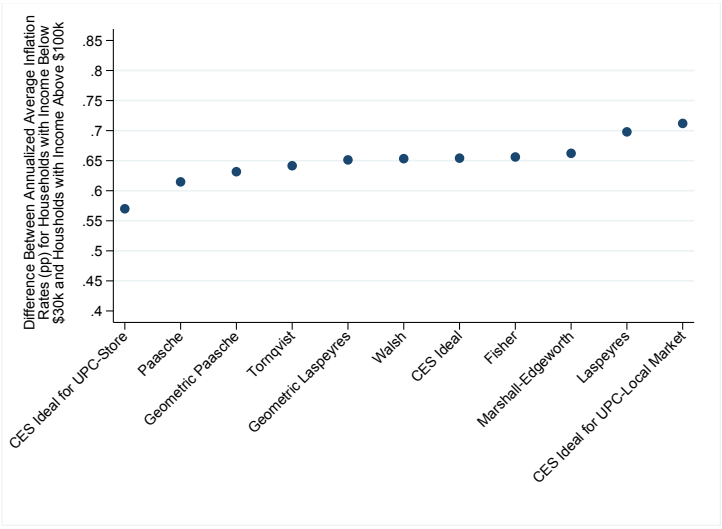
[13] I have also estimated a demand system based on the translog expenditure function, following Feenstra and Weinstein (2015). The results are qualitatively similar to those presented here and are available from the author upon request.

not affect the results). The inflation rate is about 0.65pp lower for households making more than $100,000 a year, relative to households making less than $30,000. As shown in Panel A of Figure 2, similar results are obtained when considering any of the price indices introduced in Subsection 3.2. In addition, Panel A of Figure 2 reports the inflation difference when re-defining products as UPCs available in the same store, or as UPCs available in the same local market (see Appendix B for a map of local markets). The results with this new definition of products are very similar. Overall, across all price indices and product definitions, the inflation rate is always between 0.56pp and 0.72pp lower for households making more than $100,000 a year, relative to households making less than $30,000. Panel B of Figure 2 shows that these results are robust when considering other income groups and when repeating the analysis within age groups. For each age group, inflation is sytematically lower for the higher-income households.[14]

[14] In a companion paper, Jaravel (2016) investigates patterns of inflation and product innovations across the age distribution.

Figure 2: Robustness of Inflation Difference between High- and Low-Income Households For Various Price Indices (Overlapping Products)

Panel A: For Various Price Indices



Panel B: Within Age Groups

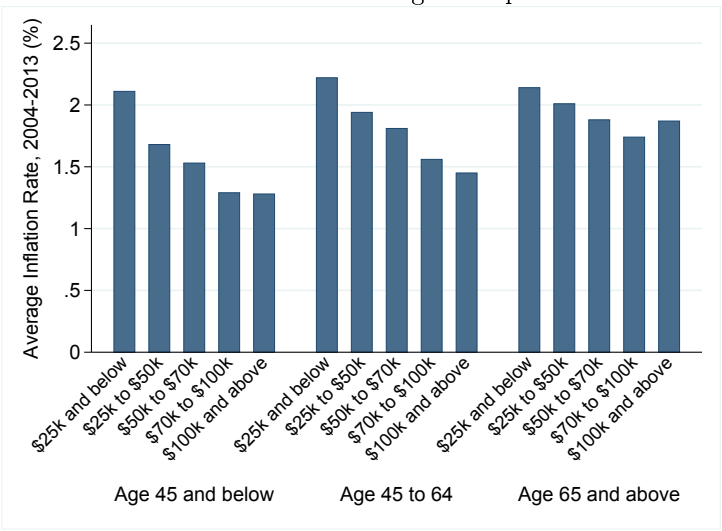


Table 1 shows the robustness of this result across subsamples. The difference between the inflation rates of high- and low-income households exists before, during and after the Great Recession,[15] and it is not driven by any single department. Appendix C presents various additional robustness tables and figures. First, Tables 11 and 12 describe the level of inflation for various cuts of the income distribution, various price indices and various periods. Figure 16 summarizes this information and shows that the difference in inflation rates is very

[15]The difference inflation rates appears to be larger during the Great Recession. Argente and Lee (2015) argue that the way in which consumers adjusted their shopping behavior to mitigate the crisis can explain the difference in the inflation rates across the income distribution during this period.

robust: higher-income households consistently experienced a lower inflation rate. Second, I redefine products to be UPCs available in a given local market (Table 13) or UPCs available in a given store (Table 14) and show that the results continue to hold. Additional robustness checks are discussed at the end of this section.

Table 1: Robustness of Inflation Difference Across the Income Distribution (Overlapping Products) For Various Periods and Departments

Period	Excluded Department	Average Annual Inflation Difference between High- and Low-Income Households
2004-2013	None	0.654
2004-2006	None	0.472
2011-2013	None	0.529
2004-2013	Health and beauty care	0.689
2004-2013	Dry grocery	0.738
2004-2013	Frozen food	0.690
2004-2013	Dairy	0.649
2004-2013	Deli	0.657
2004-2013	Packaged meat	0.654
2004-2013	Fresh produce	0.655
2004-2013	Non-food grocery	0.534
2004-2013	Alcohol	0.638
2004-2013	General merchandise	0.631

3.4.2 Decompositions

It is possible to decompose the inflation difference between households at difference points of the income distribution. For the purpose of this exercise, I focus on comparing households making more than $100,00 a year to households making less than $30,000 a year. The inflation difference reflects the combined effects of both price and quantity changes, as well as baseline differences in spending patterns across income groups. For instance, it could be that high-income households spend more on fresh produce and that inflation tends to be lower in this broad item category. Alternatively, it could be the case that high-income households experience different inflation rates compared with low-income households on the same bar codes, for instance because they shop at different stores or have different propensities to use coupons. Accordingly, the inflation difference between high income and low-income households can be decomposed into a "between" component, which corresponds to the inflation difference that would prevail if households differed only in terms of their expenditure shares and experienced the same inflation rate within an item category, and a "within" component, which corresponds to the inflation difference that would prevail if households differed only in terms of the inflation rate they experience within an item category and had the same expenditure shares across categories. Formally, for any grouping of products G, we can decompose the inflation difference between high- and low-incomme households as follows:

$$\pi^R - \pi^P \equiv \sum_G s_G^R \pi_G^R - \sum s_G^P \pi_G^P = \underbrace{\left(\sum_G s_G^R \pi_G - \sum s_G^P \pi_G \right)}_{Between} + \underbrace{\sum_G \overline{s_G}(\pi_G^R - \pi_G^P)}_{Within}$$

with s_m^i the share of spending of income group i on product grouping G andπ_G^i the inflation experienced

by income group i in product grouping G. π_G and $\overline{s_G}$ denote the average inflation rate and the average spending shares for product grouping G.

Table 2 reports the results of the decomposition at the following levels of aggregation: department, product group, product module, UPC, UPC in a given local market, and UPC in a given store. Inflation is directly observed at the product level for the last three categories, and the definitions of inflation for categories at levels of aggregation above the UPC are given in subsection 3.2. Perhaps not surprisingly, less than 10% of the differenced in the inflation rate experience by high- and low-income households is due to differences in spending across broad departments. More surprisingly, less than 25% of the inflation difference results from different spending patterns across the 125 detailed product groups, and less than 45% of the difference from spending patterns across the 1,025 very disaggregated product modules. More than 70% of the inflation difference occurs between UPCs. This is a large share of the overall difference in inflation rates, but a substantial fraction of the difference still occurs *within* UPCs. To assess the mechanism at play, I repeat the decomposition at the level of UPCs in a given local market, which brings the share of the "between" component close to 80%, as well as at the level of UPCs in a given store, which brings the share of the "between" component to 92%. Taken together, these results show that most of the difference in inflation rates between high- and low-income households occurs across UPCs, and that some of the effect results from differential price dynamics for the same UPC across stores. In Section 4, I examine whether local competition and changes in markups can explain these patterns.

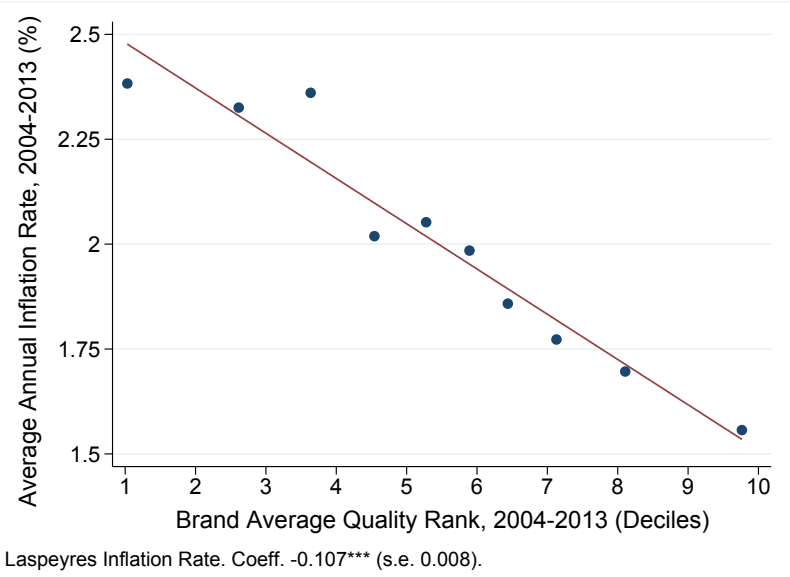
Table 2: Decomposition of the Inflation Difference Between High- and Low-Income Households

Aggregation Level (Broad to Narrow)	Decomposition	Inflation Difference	
		pp	% of actual
Department	Between	0.06	8.6
Product Group	Between	0.14	21.4
Product Module	Between	0.28	42.8
UPC	Between	0.476	72.2
UPC-Local Market	Between	0.520	78.8
UPC-Store	Between	0.607	92.1

Table 3 delves further into the differences in price dynamics across UPCs, which are of particular interest because they explain most of the inflation difference across the income distribution and because the Nielsen data is less reliable to document variation in prices paid by different income groups for the same UPC. Indeed, Nielsen often automatically enters the price of the UPC based on the store the panelist reported for their

shopping trip.[16] Figure 3 documents that within product modules inflation is lower for products that belong to brands in higher price deciles. The price deciles are computed within each module based on the average (spending-weighted) unit price of the products that belong to a brand, over all years in the dataset. This approach provides a way to segment the product space even within product modules, the highest level of disaggregation provided by Nielsen. It is not subject to mean reversion because the deciles are not based on the price of the UPC itself, but rather on pricing behavior at the brand level over the entire dataset. Table 3 shows that differences in the spending patterns of high- and low-income households across price deciles within product modules explain more than 85% of the inflation difference between high and low-income households that exists across UPCs. Taken together, the decomposition exercises show that the inflation difference between high- and low- income households primarily exists across UPCs, rather than within, and that it can be accounted for by the fact that inflation is lower for products belonging to higher-quality brands, which primarily cater to higher-income consumers.

Figure 3: Inflation across Brand Price Deciles within Modules



[16] Nielsen obtains the price data from that store in that particular week.

Table 3: Decomposition of the Inflation Difference Between High- and Low-Income Households Relative to Across-UPC Benchmark

Aggregation Level (Broad to Narrow)	Decomposition	Inflation Difference	
		pp	% of benchmark
Department	Between	0.061	**12.8**
Product Group	Between	0.143	**30.0**
Product Module	Between	0.282	**59.2**
Product Module*Price Decile	Between	0.408	**85.7**
UPC	Between	0.476	**100**

3.4.3 Relevance for the Methodology of Statistical Agencies

Table 2 means that product-level data is needed to capture the magnitude of the difference in inflation rates between households at different points of the income distribution. It is not sufficient to simply reweight aggregate price series based on income-specific spending shares, even when the level of aggregation is as detailed as product modules. Yet this is precisely the approach followed by the BLS and other statistical agencies. More specifically, the BLS collects prices on 305 different item categories, known as "entry-level items" (ELI). Most of these item categories are very coarse. 230 of them are actually in the retail sector, where the level of disaggregation is much higher than in other sectors. Still, this level of aggregation is too high to capture the bulk of the difference between high and low income consumers. This explains why the result presented here may appear inconsistent with the existing literature, which has found small differences between high and low income consumers.

For instance, McGranahan and Paulson (2005) compute income-specific inflation rate based on between-ELI inflation differences and income-specific CEX spending patterns. Using their data, I computed that between 2004 and 2013 the annualized inflation difference for households in the bottom vs. top income quartiles was 0.18pp, which is similar to what I obtained in the Nielsen data with the "between product group" methodology (see Appendix C for details).

Therefore, the conventional wisdom that inflation is similar across the income distribution may be misplaced: statistical agencies like BLS collect data at a broad level of aggregation, which biases the estimate of the difference in inflation across income groups towards zero. Using the Nielsen data, I directly show that the magnitude of this bias is large in the retail sector. Table 3 suggests that a large share of the inflation difference across income groups could be captured by segmenting each of the detailed item categories by price deciles - the data collected by statistical agencies like the BLS could be used to replicate this approach.[17]

[17] One would then need to infer the spending shares of various income groups along price deciles, which could be done for instance by estimating "quality Engel curves" as in Bils and Klenow (2001).

3.4.4 Related Literature

My results are consistent with Argente and Lee (2015), who study the inflation difference for high- and low-income households during the Great Recession, find that it is lower for higher-income households and argue that this effect is driven by substitution patterns. The inflation dynamics I describe in this paper are more general and of a different nature: I show that the difference in inflation rates across the income distribution extends well beyond the crisis and continues to hold even when substitution effects are ignored (indeed, Figure 2 shows that the magnitude of the inflation difference is similar across a variety of price indices that do not allow for substitution, like the Laspeyres index). In Section 4, I show that the magnitude of the inflation difference between high- and low-income households can be explained by the equilibrium response of supply to market size effects.[18]

Two other recent papers are closely related to my findings. Pisano and Stella (2015) document that lower-income households pay lower prices than higher-income households for the same products, primarily because they shop more at discount stores. In contrast, I focus on changes in income-specific price indices over time and use the demand system to provide a measure of quality-adjusted inflation. Faber and Fally (2015) explore the implications of firm heterogeneity for household price indices across the income distribution. They find that larger, more productive firms endogenously sort into catering to the taste of wealthier households, and that this gives rise to asymmetric effects on household price indices in their structural model. I provide direct evidence of differences in inflation rates across the income distribution and, in Section 4, I focus on a completely different explanatory mechanism.

To the best knowledge, my paper is the first to propose decompositions of the inflation differential between high- and low-income households as in Section 4.3.2. and to relate these patterns to the dynamics of product creation and endogenous changes in markups, which are discussed in the remainder of the paper. My analysis shows that collecting product-level data is key to accurately measure the divergence of inflation rates across the income distribution - this methodological lesson is likely to apply to other sectors beyond retail. The sign and magnitude of the inflation difference between high- and low-income households in other sectors remains an open question.[19]

3.5 Changes in Product Variety Across the Income Distribution

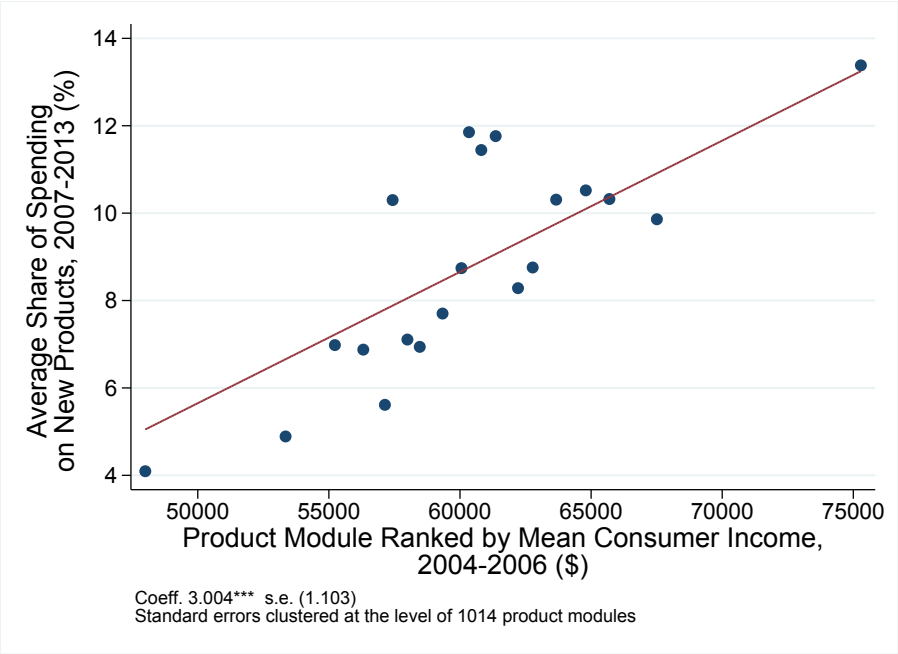
Do welfare effects from increasing/decreasing product varieties also differ across the income distribution? I find that the rate of increase in product variety is faster in product modules catering to higher-income households, implying that higher-income households benefit more from increasing product variety. Figure 4 shows this effect by using the share of spending on new products (barcodes which did not exist in the

[18] Note that both my results and the results of Argente and Lee (2015) appear inconsistent with the findings of Broda and Romalis (2009), who also use Nielsen data and report in an unpublished manuscript that they find that inflation is lower for lower-income households.

[19] See the debate between Moretti (2013) and Diamond (2016) about price changes in the housing sector for college and high-school graduates. To my knowledge, there have been very few attempts at developing non-homothetic price indices and McGranahan and Paulson (2005) remains the main reference. Handbury (2013) developed a non-homothetic price indices across cities but did not have panel data to study inflation dynamics.

previous year) as a measure of the flow of successful product innovations. For every $10,000 increase in the mean income of the consumers buying from a product module, the share of spending in this product module goes up by 3 percentage points, about a third of the average share of spending on new products. Appendix D reports patterns of product destruction across product modules, which are much more homogeneous.

Figure 4: Product Variety Increases Faster In Product Modules Catering to Higher-Income Households



According to the demand system presented in Section 3, the share of spending on new products is the welfare-relevant metric. However, is supply or demand driving the relationship shown on Figure 4? It could be the case that more products are introduced in product modules catering to high-income consumers because of supply effects, which may be exogenous (e.g. it may be inherently easier to introduce new product at the high-end of the product space) or endogenous (e.g. if innovators and suppliers decide to specifically target higher-income consumers). Alternatively, it could be the case that higher-income consumers have a higher taste for novelty and purchase new products wherever they are introduced in the product space. In other words, the share of spending on new products may be higher in product modules catering to higher-income households simply because new products diffuse faster (while the rate of product introduction may be similar across modules). To isolate the contribution of supply, the ideal regression would compare the same household moving across the product space. Such a regression can be directly run in the Nielsen data, at the household $H \times$ product module M level with household fixed effects:

$$ShareSpendingNewProducts_{HM}$$

$$= \alpha + \beta ProductModuleIncomeRank_M + \alpha_H + \epsilon_{HM}$$

where α_H is a household fixed effect and $ProductModuleIncomeRank_M$ is the rank of the product module by income of the representative consumer in the product module (computed using 2004-2006 data). The results are reported in Table 4, with standard errors clustered at the household level. As in the previous graphs, I find a strong positive relationship between the share of spending on new products and the mean income of the consumer in the product module. This analysis confirms that supply plays a role in this process, because household fixed effects ensure that the relationship is not driven by a composition effect across modules (i.e. different propensities of consumers to buy new products wherever they show up in the product space). I also present specifications with interaction terms for whether the household is "high-income" (income above $100,000) or "low-income" (income below $30,000). The magnitude of the interaction effects is small, around 10% of the effect for middle-income households.

Table 4: New Products Target Higher-Income Consumers

	$ShareSpendingNewProducts_{hm}$	
$ProductModuleIncomeRank_M$	2.79***	2.82***
	(1.024)	(1.031)
$ProductModuleIncomeRank_M \times HighIncome_H$		-0.24***
		(0.063)
$ProductModuleIncomeRank_M \times LowIncome_H$		0.11*
		(0.058)
$Household\ Fixed\ Effects$	Yes	Yes

Standard errors clustered by product modules.

I have checked that similar results hold for other measures of "new products" - new UPCs relative to two, three or four years ago, as well as new brands. Appendix D reports additional results. In particular, Figure 21 shows that, across product modules, the rate of increase in the total number of varieties increases by one percentage point with a $10,000 dollar increase in the income of the representative consumer. I have also investigated patterns of product destruction, which are much more homogeneous across the product space. Moreover, I have examined patterns of product creation and destruction within product modules, which are similar to the patterns observed across modules (i.e. even within a module, higher income households tend to benefit more from product creation and destruction). Table 19 shows that the differences in shares of spending on new products between high- and low-income consumers largely occur within product modules - this pattern is very similar to the inflation decomposition discussed earlier and provides preliminary evidence that there is a tight connection between the inflation and product innovation patterns, which is further examined in Section 4.

3.6 Quality-Adjusted Inflation Across the Income Distribution

Using the results from equation 2, I can bring together the previous facts on inflation for products available in consecutive years and patterns of product creation and destruction. Table 5 shows the distribution of the estimated elasticities of substitution by income groups. Two findings stand out. First, the elasticities tend to be slightly smaller for higher-income households, i.e. higher-income households are less price elastic in equilibrium.[20] This provides direct evidence of non-homotheticities (in addition, Figure 22 in Appendix C shows that the elasticites for high- and low-income consumers are not very strongly correlated). Second, the magnitude of the elasticities is very high. Using the optimal markup formula derived in Section 2, these magnitudes are consistent with the observed markups in the retail sector.[21] The high values of the elasticities means that the "product variety" adjustment is very small: since the elasticities of substitution are very high, most of the welfare effects are captured by the inflation difference on goods that exist across consecutive years (cf. derivation in subsection 4.2). As a result, the quality-adjusted inflation across the income distribution (Figure 23 in Appendix C) looks virtually identical to inflation across the income distribution for overlapping products (Figure 1).

Table 5: Distribution of Estimated Module-Level Elasticites of Substitution For Three Income Groups

		Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Percentile	10th	9.08	9.30	8.56
	25th	13.91	13.44	13.47
	50th	21.45	19.87	19.78
	75th	35.30	32.74	34.25
	90th	65.98	56.75	71.27

3.7 Further Robustness Checks

Selection effects. A potential concern is that the inflation patterns described above could result from selection effects, for instance if low-income households overwhelmingly consume goods whose characteristics are rendered obsolete by the entry of new products. In such a case, a relatively higher share of the goods consumed by the poor would be exiting the market in any given year - the price changes for these goods are not observed, but if they were they would be negative because these products face tougher competition.[22] Tables 16 to 18 show that such selection effects are in fact not at play in the data.[23]

Price Convergence Anoter potential concern is that the observed inflation difference between high- and low-income households could be driven by the fact that high-income households might initially pay a higher

[20] Note that the equilibrium elasticity of susbtitution depends on consumers' preference parameters, but also on the competitive environment if the elasticity of substitution is not constant. See Section 3 for a discussion of models of monopolistic competition with variable elasticities of substitution. I have estimated a demand system based on the translog expenditure function, which features decreasing elasticities of substitution, and have obtained qualitatively similar results.

[21] In retail (groceries and food) the margin is around 2.71%.
See http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/margin.html

[22] See Pakes and Erickson (2011) for a discussion of such selection effects.

[23] Note that even if selection effects were at play, the nested CES structural demand system with new goods addresses these concerns by adjusting the price index when new varieties enter.

price for the same UPCs as low-income households, and the price would then converge to the same level for all households in future periods. The last three rows of Table 2 reject the hypothesis by showing that the "within-UPC" share of the total inflation difference is modest. A more direct way of showing that this mechanism is not the driving force, without the need for any assumption about the demand system, is to run a regression of the unit price of the UPC on a UPC fixed effect and an indicator for whether the household is high income (restricting attention to products purchased by both income groups). Table 6 reports the results of such a regression and show that, in any given year, households making more than $100,000 a year tend to pay about 2.9% more for the same UPC, compared with households making less than $30,000 a year. This result is consistent with the findings of Pisano and Stella (2015). The magnitude of this effect is negligible compared with the 0.65pp difference in inflation rates, which over the course of a few years leads to a much bigger welfare difference between high- and low-income households than the difference in price levels in any given year.[24] Figure 17 in Appendix C provides complementary evidence by showing that the distribution of average unit prices paid by high- and low-income households is very similar, restricting attention to the set of products purchased by both income groups.

Table 6: Differences in Price Level Paid for Same UPC by High- and Low-Income Households ($)

	Average Unit Price
High-Income Household	0.0664***
	(0.00118)
Constant	2.2825***
	(0.00061)
UPC*Year Fixed Effect	Yes
R^2	0.9954

The product cycle. One may worry that the patterns about inflation and new products are driven by the "product cycle" - namely, products start in the market with a very high price, and at that point are only purchased by high-income households, and then converge to their long-run, stable price, at which point they start being purchased by lower-income households. I address this concern in several ways. First, my results hold across the product space, as shown in Figures 4 and 9. If the product cycle was driving the results, then the measured differences in inflation and product innovation should only be visible from the point of view of each individual consumer and not across the product space. Second, I have repeated the analysis by considering only products in the middle of their lifecycle. Specifically, in any given year I have resticted the sample to products that had been in the market for at least two years and that would remain in the market for at least two more years. The inflation patterns obtained with this approach are similar to those reported above. Third, I have shown that the product cycle is not an important force in the data as barcodes do not travel down the income distribution (empirically, barcodes tend to remain in the same price decile during

[24] Note that my focus on inflation allows me to take into account changes in product variety and consumer substitution across products over time, as well as to characterize how these patterns differ across the income distribution. The static analysis of the levels of prices paid for the same barcodes by individuals across the income distribution does not speak to these dynamic consideration, which are first order in the data.

their entire lifecycle, which is intuitive for the retail sector and stands in contrast with other products like computers). Fourth, even if the product cycle was an important force in the data, under the assumption described at the beginning of this section the nested CES demand system will provide an accurate estimate of the quality-adjusted inflation rate for each of the various income groups, given the speed of the product cycle. In particular, in this analysis the "novelty" of a product is determined separately for each income group based on the basket of goods consumed by this income group in the previous year.

The fashion cycle. A distinct concern is that the inflation patterns may be driven by a phenomenon analogous to the "fashion cycle" - the fact that products exhibit seasonality patterns and that the price of older products falls disproportionately. For instance, because of the fashion cycle measured inflation is negative in the apparel industry - yet productivity gains for apparel are small and it would be incorrect to infer large welfare gains from the observed price patterns.[25] Conceptually, the fashion cycle means that the assumption that the "quality" of a barcode is fixed over time fails - if newness is a key feature of the utility derived from a product, the observed price of this product will fall over time but this may not reflect any change in the quality-adjusted price. I address the concern that high-income households may be more like to be affected by the fashion cycle in two ways. First, the fashion cycle is about churn of products and not about a net increase in the number of available varieties. I show that there is a faster increase in varieties in the parts of the product space that cater to higher income households, but there is not more churn. Similarly, the price patterns across product modules are predicted by the net increase in product variety, rather than by churn. Second, the results hold even with product categories where the fashion cycle is unlikely to exist, such as food product .

Local market effects and store effects. Using a reweighting procedure, I show that the differences in spending patterns across stores and local markets cannot explain the inflation patterns previously documented. The results are reported in Appendix Table 20.

Alternative measures of household income. I repeated the analysis with three alternative measures of household income: reported income divided by household size; total retail expenditures per capita within a household; and whether the head of household is a college graduate.

Sampling variability. To ensure that the results are not driven by differing degrees of sampling variability across income groups, I built a random subsample of the data with an equal number of households in each of the income bins. I have also checked that the results across product modules hold in the Retail Scanner Data (which is based on information recorded directly at the store, not obtained from households, and contains many more observations as described in Appendix B).

Base drift. I have repeated this analysis using unchained price indices instead of chained indices and obtained similar results.

Quarterly data. Table 15 shows that the results are very similar when repeating the analysis at a quarterly frequency.

[25] The Bureau of Labor Statistics addresses this by making hedonic adjustments and by ignoring sale prices.

4 The Equilibrium Response of Supply to Changes in Demand

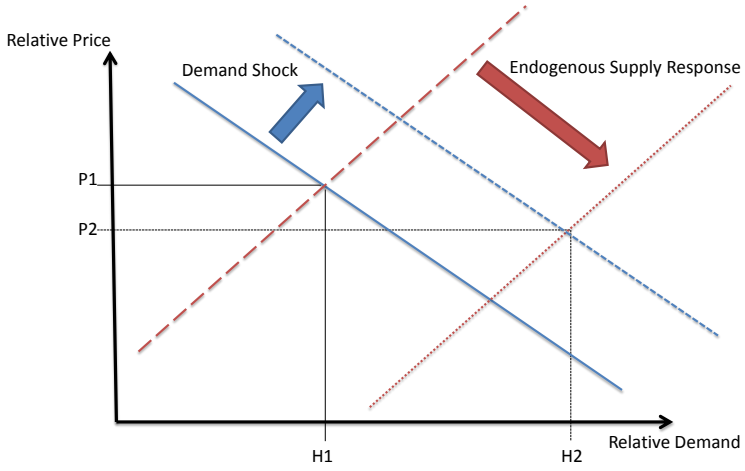
In this section, I investigate the hypothesis that an important driver of the results presented in Section 3 - namely, the fact that higher-income household experience lower inflation and a faster increase in product variety from 2004 to 2013 - is differential income growth and directed product innovations. I first present a theory showing how rising nominal income inequality may have a causal effect on the direction of product innovations, and in turn result in a further increase in real income inequality. An increase in market size leads to more product entry, which puts downward pressure on the prices of existing products through increased competition. I test the key channels of this theory, first by studying a series of descriptive patterns on price and markup dynamics and then by tracing out the causal impact of a demand shock on price and innovation dynamics. I then discuss other possible mechanisms.

4.1 Theory

4.1.1 Intuition: Tracing Out the Observed Long-Term Supply Curve

Because of nonhomothetic preferences and the endogenous price changes induced by changes in relative demand, changes in nominal inequality may overstates or understate changes in real inequality. Consider Figure 5. When relative demand goes up, if the short-run supply curve is upward sloping as in standard price theory, then the equilibrium price should go up. However, supply may endogenously shift out due to the response of firms to market size effects, which will at least mitigate the price increase and, as illustrated in Figure 5, could potentially result in a new equilibrium price that is lower than the initial equilibrium price. This "price overshooting" case is shown to be the empirically-relevant case in Section 4. In other words, the observed long-term supply curve is downward sloping.[26]

Figure 5: Does the Price Fall When Demand Rises?



[26] The observed long-term supply curve is defined as the nexus of equilibrium points traced out by shifts in the demand curve.

To investigate whether changes in nominal inequality overstate or understate changes in real inequality, the following concepts are useful:

- **Weak equilibrium (relative) bias** ("directed technical change"): when demand for a good becomes relatively more abundant, supply (technology, innovation, entrepreneurship, etc.) becomes endogenously biased towards this factor.
- **Strong equilibrium (relative) bias**: the relative supply curves for goods are downward sloping.

Consider demand H for a high-quality good and demand L for a low-quality good. Endogenous technology A is a function of relative demand $\frac{H}{L}$. The equilibrium relative price is

$$\frac{p_H}{p_L} = f\left(\frac{H}{L}, A(\frac{H}{L})\right)$$

There is weak equilibrium bias if:

$$\frac{\partial f}{\partial A}\frac{\partial A}{\partial H} < 0$$

There is strong equilibrium bias if:

$$\frac{\partial f}{\partial H} + \frac{\partial f}{\partial A}\frac{\partial A}{\partial H} < 0$$

where $\frac{\partial f}{\partial H} > 0$, as in standard price theory.

The equations above and 5 provide an intuitive reduced-form way of thinking about the effect of shifts in demand on the equilibrium price. In the next subsection, I discuss a specific microfounded model that is consistent with the evidence, generates additional predictions and provides a framework for estimation and for welfare calculations.

4.1.2 A Microfoundation

I focus on microfounded models of monopolistic competition with free entry of products. This broad class of models is appealing for three reasons: the assumption of monopolistic competition is reasonable in retail (the Herfindahl index for most product groups is below 0.20), these models nest the standard model of directed technical change (Acemoglu, 2002), and they generate rich product-level predictions and counterfactuals. The intuition for the effect of changes in market size on supply in monopolistic competition models is as follows: an increase in market size leads to more product entry, which puts downward pressure on the prices of existing products (pecuinary externality). Therefore, in such models innovation occurs entirely through product entry - there is no "process innovation" reducing the marginal cost of the existing products, whose price dynamics are determined by changes in markups.

Within the class of monopolistic competition models with free entry of products, only some models are consistent with the "price overshooting" case illustrated in Figure 5. In particular, the CES model of Acemoglu (2002) does not allow for the possibility that the price goes down when demand goes up (see Appendix A

for a detailed derivation). On the other hand, Melitz and Ottaviano (2008) is consistent with the strong equilibrium bias (see Appendix A for a derivation). In the rest of this section, I characterize the conditions under which "price overshooting" is possible using the general monopolistic competition model of Zhelobodko, Kokovin, Parenting and Thisse (2012). The key insight is that, in general equilibrium, the curvature of the utility function and variable markups drive the sign and magnitude of the response of the equilibrium price to changes in market size.

L consumers with additively separable preferences over varieties solve:

$$max_{x_i \geq 0} \, U = \int_0^N u(x_i)di \qquad s.t. \qquad \int_0^N p_i x_i di = E$$

Consumer maximization yields

$$p_i(x_i) = \frac{u'(x_i)}{\lambda}$$
$$\lambda = \frac{\int_0^N x_i u^{'}(x_i)di}{E}$$

Total quantity demanded is $q_i = Lx_i$. The monopolist takes the residual demand curve as given and solves:

$$max \; \pi(q_i) = R(q_i) - C(q_i) \equiv \frac{u'(q_i/L)}{\lambda} q_i - V(q_i) - F$$

with $V(.)$ is the variable cost function and F the fixed cost. The optimal markup of the producer is therefore given by:

$$M^* = -\frac{x_i \cdot u''(x_i)}{u'(x_i)}$$

At the free entry equilibrium, $\pi(q_i^*) = 0$ and a mass N^* of firms satisfies labor market clearing:[27]

$$N^* = \frac{L \cdot E}{C(q_i^*)}$$

Therefore, the model delivers the following comparative statics:

$$\frac{dN^*}{dL} > 0 \qquad \frac{dx_i^*}{dL} < 0 \qquad \frac{dM_i^*}{dL} \gtreqless 0$$

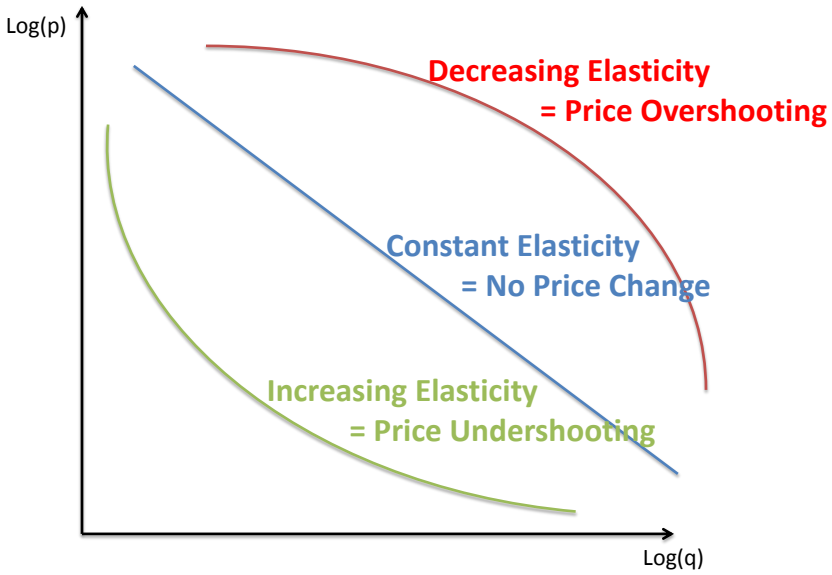
Therefore, the optimal markup is given by the inverse of the price elasticity of demand[28] (which, given our assumption of separability, is equal to the inverse of the elasticity of substitution between varieties). This result is very general and holds regardless of the shape of the cost function $V(.)$. It shows why the equilibrium response of prices to changes in market size crucially depends on the curvature of the utility function. The intuition for the comparative statics is as follows. When market size increases, new products enter the market. As a result, consumers start spreading out their expenditures across more products, due to taste for variety. Consequently, consumption per capita x_i for the existing products goes down, which induces a responses of the optimal markup M^*. The equilibrium markup which may increase, decrease or

[27] A similar model can be solved by assuming that the sector is small relative to the total economy, which allows for ignoring some GE effects. See Mayer, Melitz and Ottaviano (2016).

[28] This term is also equal to the coefficient of relative risk aversion.

stay unchanged, depending on the properties of demand. Figure 6 shows this effect in log-log space. The blue curve corresponds to CES demand, as in Acemoglu (2002). Movements along the curve do not matter, the elasticity is constant. On ther other hand, the red curve shows that when consumption per capita decreases (moving to the left along the curve), the price elasticity of demand goes up, i.e. the optimal markup goes down. Melitz and Ottaviano (2008) corresponds to this case. Conversely, as shown with the green curve, if the price elasticity of demand is increasing the equilibrium price should go up in response to an increase in market size.

Figure 6: The Equilibrium Response of Price to Changes in Market Size Depends on the Price Elasticity of Demand



Because preferences are nonhomothetic, the curvature of the utility function may differ for consumers in different income groups and the equilibrium response of price to market size may be different in product modules catering to different consumer segments. This framework allows for rich counterfactuals to answer the question: what would have been the difference in inflation and rates of product introduction across the income distribution absent the endogenous response of supply to market size effects? The framework is based on homothetic utility functions within product modules, but I can separately estimate these utility functions for different groups of consumers across the income distribution, which effectively allows for nonhomotheticities by letting the parameters of the utility function vary freely with the level of income.[29]

[29]Section 3 shows how to nest the various sub-utility functions for each product module into one aggregate utility function.

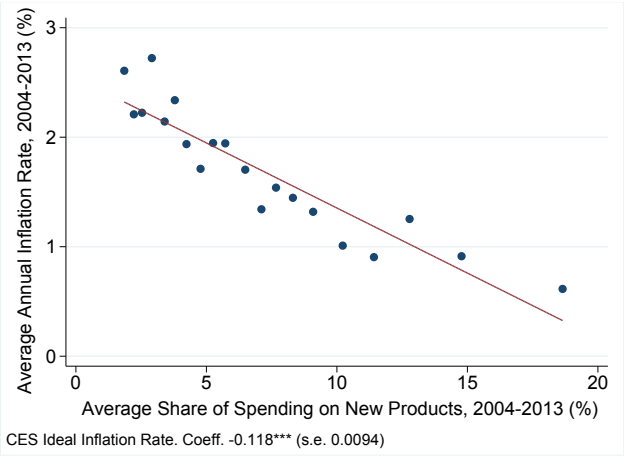
In sum, the model makes three predictions: 1. there is a strong negative correlation between inflation and the share of spending on new products both *across* and *within* product modules; 2. the inflation patterns across the income distribution are driven by differences in changes in markups; 3. growing demand causes more product innovations and lower inflation. I test and find support for all three predictions in the rest of this section.[30]

4.2 The Relationship Between Inflation and Share of Spending on New Products

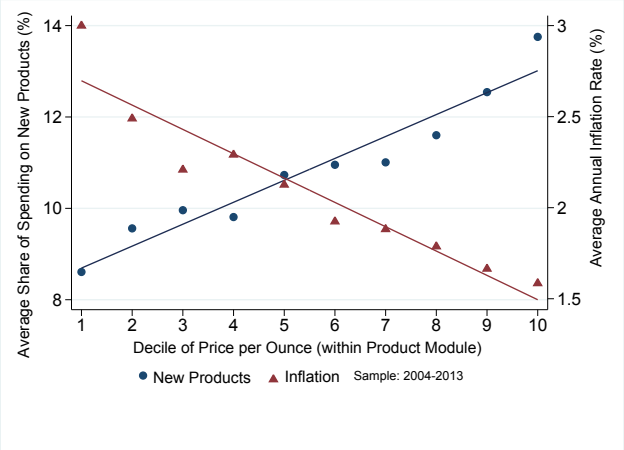
The negative correlation between inflation and the share of spending on new products is a key feature of the data. Panel A of Figure 7 shows this relationship across product modules. Panel B shows that the relationship persists within product modules when segmenting the product space by price deciles: product entry is higher in higher price deciles, while inflation is lower.

Figure 7: The Negative Relationship Between Inflation and Share of Spending on New Products

Panel A: Across Product Modules



Panel B: Within Product Modules



[30] Note that these tests speak to an active debate in the trade literature about the source of the gains from trade and the role of variable markups and variable elasticity of substitution preferences. See in particular DeLoecker, Goldberg, Pavcnik and Khandelwal (2012), Feenstra and Weinstein (2016), and Mayer, Melitz and Ottaviano (2016)

A simple decomposition exercise shows that the relationship between inflation and product innovations across modules can explain a large fraction of the inflation patterns across income groups documented in Section 3.[31] As previously mentioned, for any product grouping G, we can write the inflation difference between income groups as:

$$\pi^R - \pi^P \equiv \sum_G s_G^R \pi_G^R - \sum s_G^P \pi_G^P = \underbrace{\left(\sum_G s_G^R \pi_G - \sum s_G^P \pi_G \right)}_{Between} + \underbrace{\sum_G \overline{s_G}(\pi_G^R - \pi_G^P)}_{Within}$$

with s_m^i the share of spending of income group i on product grouping G,π_G^i the inflation experienced by income group i in product grouping G, and with π_G and $\overline{s_G}$ denoting the average inflation rate and the average spending shares for product grouping G. We can now decompose the "between" component further and examine how much of the inflation difference across categories is explained by (or predicted by) differences in shares of spending on new products across categories:

$$\left(\sum_G s_G^R \pi_G - \sum s_G^P \pi_G \right) = \left(\hat{\pi_G^R} - \hat{\pi_G^P} \right) + R$$

with

$$\begin{aligned} \hat{\pi_G^R} - \hat{\pi_G^P} &= \sum_G \hat{\beta} X_G (s_G^R - s_G^P) \\ R &= \sum_G \hat{\epsilon_G} (s_G^R - s_G^P) \\ \pi_G &= \beta X_G + \epsilon_G \end{aligned}$$

where X_G is share of spending on new products in G. $\hat{\beta}$ is the OLS estimate of β. This procedure calibrates the extent to which the difference in inflation rates between high- and low-income households results from the fact that high-income consumers tend to devote a higher share of their spending to product categories where the rate of product innovations is higher (i.e. moving to the right along the x-axis in panel A of Figure 7), or from the fact that high-income households tend to spend more on product categories with a lower share of inflation, holding the rate of product innovations constant (i.e. moving down the y-axis in panel B of Figure 7). Table 7 shows that for the various levels of aggregation, around half of the inflation difference between high- and low-income households can be explained by differences in patterns of product innovations.[32] These results provide strong support for the first prediction of the model and the notion that the joint dynamics of product innovation and inflation are crucial to understand changes in real inequality.

[31] This is similar in spirit to the reweighting technique introduced in DiNardo, Fortin and Lemieux (1996).

[32] Note that any measurement error (e.g. UPC relabeling that does not reflect a true product innovation) will bias this esitmate downward, therefore these estimates can be viewed as a lower bound.

Table 7: How Much of the the Difference in Inflation Between High and Poor is Explained by Patterns of Product Innovations?

Aggregation Level (Broad to Narrow)	Share of Rich-Poor Inflation Difference Explained
Department	40.9
Product Group	58.3
Product Module	51.3

The relationships described so far are only correlations and should not be interpreted as causal, but they provide transparent evidence on the pervasive nature of the relationship between inflation and product innovations and on its relevance for understanding changes in real inequality.

4.3 The Role of Markups

As mentioned in Section 2, I observe retailer price p_{it} and wholesale cost c_{it} from 2004 to 2007 for a subset of the product. A first-order Taylor expansion yields a convenient additive expression for the log price change:

$$p_{it} = m_{it} + c_{it}$$
$$\Delta^t log(p_{it}) \approx \Delta^t log(c_{it}) + \Delta^t \frac{m_{it}}{c_{it}}$$

I can then run the following regression across product modules, with store-year fixed effects to absorb rent and labor costs:

$$\Delta^t log(p_{it}) = \beta I_i + \lambda_{st} + \epsilon_{it}$$
$$\Delta^t log(c_{it}) = \widetilde{\beta} I_i + \widetilde{\lambda_{st}} + \widetilde{\epsilon_{it}}$$
$$\Delta^t \frac{m_{it}}{c_{it}} = \bar{\beta} I_i + \bar{\lambda_{st}} + \bar{\epsilon_{it}}$$

with I_i income rank of module. Note that $\beta \approx \widetilde{\beta} + \bar{\beta}$. We can use this relationship[33] to answer the following question: do prices rise slower for high-income consumers because retailer margins decline faster or because wholesale costs rise slower?

As shown in Figure 8 and Table 8, changes in retailer margins account for 57% of the differential inflation between high- and low-income households. This number can be thought of as a lower bound on the total share of changes in markups in the overall inflation difference, because wholesalers' themselves have a markup. I have checked that this relationship is robust across years and with using other specifications. These results provide strong support for the second prediction of the model: variable markups are a key channel.[34]

[33] As can be checked from the regression table, the margins are sufficiently small for the Taylor expansion to be almost exact, which in turn implies the teh relationship between the regression coefficients is almost exact.

[34] Variable markups are often studied in the macro literature in the context of short-run business cycle fluctuations. The fact that markups explain a large share of the difference in inflation between high- and low-income households does not mean that

Figure 8: Changes in Wholesale Costs vs. Changes in Retailer Margins

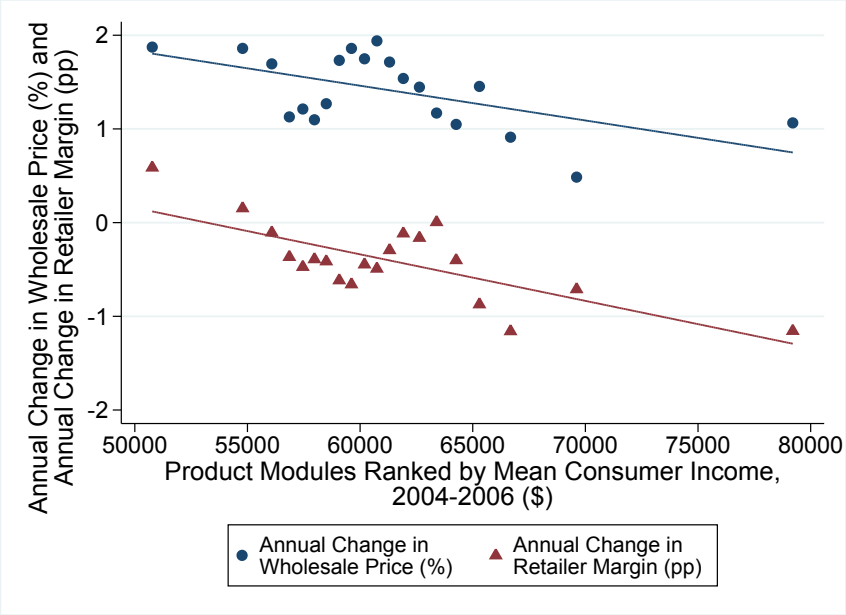


Table 8: Changes in Wholesale Costs vs. Changes in Retailer Margins

	Log Price Change	Log Wholesale Cost Change	Retailer Margin Change (pp)
$ProductModuleIncomeRank_M$	-0.777***	-0.341***	-0.448***
	(0.188)	(0.103)	(0.212)
Spending Weights	Yes	Yes	Yes
Store-Year Fixed Effects	Yes	Yes	Yes
Number of Observations	6,002,235	6,002,235	6,002,235
Number of Clusters	628	628	628

Standard errors clustered by product modules.

4.4 The Causal Effect of Growing Demand on Product Innovations and Inflation

4.4.1 Motivating Evidence and Identification Challenge

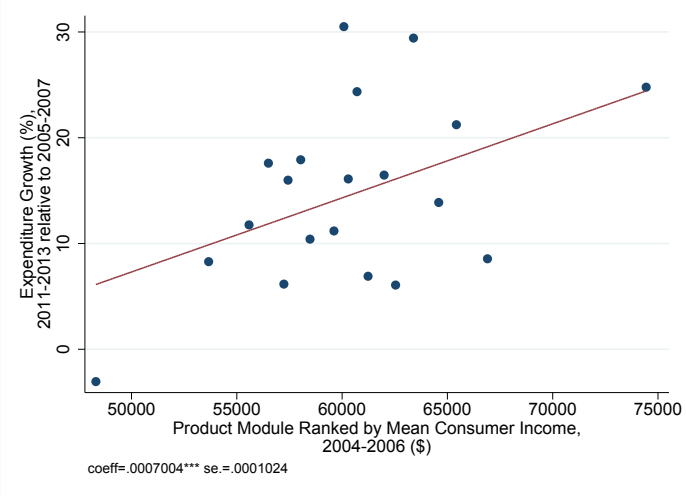
They key causal channel in the model is that growing demand causes more product entry, and in turn lower prices on existing products due to a fall in markups. Figure 9 shows that product modules catering to higher-income households indeed have both higher growth and lower inflation. Moreover, Table 4 provided early evidence that supply factors play an important role in differential product introductions across the product space.

these dynamics are bound to be short lived. Indeed, the set of available products changes over time. Adjusted for quality, the marginal cost of the new products is lower than that of existing products, which are forced to reduce their markups. In other words, the price effects show up largely through changes in markups, but these changes reflect the productivity gains brought about by new products.

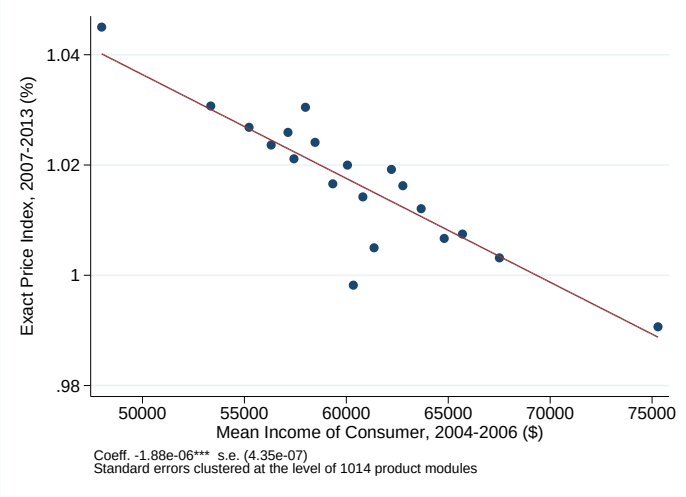
However, these facts alone do not establish that the endogenous response of supply to demand is a key channel. The equilibrium relationship between price and quantity across product modules modules does not identify the causal effect of demand, because of reverse causality (demand might be following supply) and omitted variable bias (there might be unobserved heterogeneity in the difficulty of innovating across modules, which could happen to coincide with spending patterns from non-homothetic preferences). In the remainder of this section, I build a predictor of (potential) demand that is plausibly orthogonal to supply factors. Specifically, I consider changes in market size across product modules over time at the national level driven by changes in the age and income distributions. In robusness checks, I use variation in market size both over time and across local markets within the US.

Figure 9: Product Modules Catering to Higher Income Households Have Faster Growth and Lower Inflation

Panel A: Growth Across Product Modules



Panel B: Inflation Across product Modules



4.4.2 Research Design

A major difficulty in any investigation of the impact of market size on innovation is the endogeneity of market size: better products will have larger markets. A strategy to overcome this problem is to exploit variations in market size driven by US demographic changes, which should be exogenous to other, for example scientific, determinants of innovation and entry of new products. To estimate potential market size, I construct age-income profiles of users for each product module $\times$ price decile, and then compute the implied market size from aggregate demographic changes given these (time invariant) income-age profiles. This identification strategy is similar to Acemoglu Linn (2004). Using this strategy, Acemoglu and Linn (2004) showed that large R&D efforts in the pharmaceutical industry endogenously respond to market size. By focusing on product innovations in retail, I study innovation dynamics of a very different nature. More importantly, this paper is the first to examine the causal effect of changes in market size on the price of existing products, as well as on the aggregate price taking into account the welfare gains from increased product variety. I find that prices go down when demand goes up, i.e. the observed supply curve is downward sloping.[35]

The predictor of market size is built as follows. At the beginning of the sample (2004-2006), I compute per capita expenditures $E_{MG}^{T_0}$ in product module $\times$ price decile M for fifteen age-income groups G I consider.[36] Then, I predict (potential) demand at time t as:

$$D_{Mt} = \sum_{G} E_{MG}^{T_0} P_{Gt}$$

Thus, the spending profiles are kept consant and the variation in predicted demand comes entirely from changes in age-income group size P_{Gt}. To implement this design, I compute growth of demand based on the change in the size of each age-income group in 2011-2013 relative to 2004-2006.

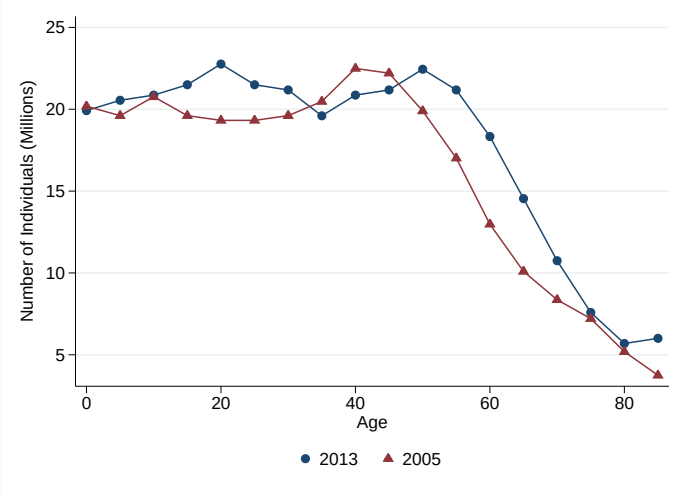
The identification assumption is that the direct effect of changes in age-income group size on the equilibrium price was only through demand. For instance, if the 20-year old were better at creating new products targeting people in the same age group, the identification assumption would be violated. As a robustness check, I repeat the analysis for older households, who are closer to retirement age. I find similar point estimates, which suggests that direct supply effects are not driving the results.

[35] In ongoing work, I use markups as another outcome to test additional predictions of the model presented at the beginning of this section.

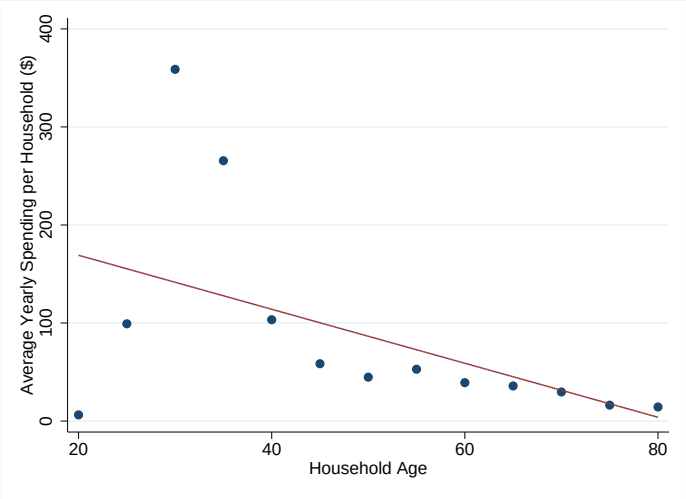
[36] Specifically, I consider the interaction of three age groups - below 45, between 45 and 65, and above 65 - and five income groups - annual household income below 25k, 25k to 45k, 45k to 60k, 60k to 100k, and above 100k.

Figure 10: Changes in Market Size from Changes in the Age Distribution

Panel A: Changes in Age Distribution, 2005 to 2013



Panel B: Example of a Product with a Steep Age Profile, Diapers



4.4.3 Results

I first present the results with a series of binned scatter plots, where each dot represent 10% of the data. I then show the results in a regression table, with standard errors clustered by product module. Figures 11 and 12 below show that the predicted increase in market size (based on the changes in the age and income distributions) is positively correlated with the introduction of new products and negatively correlated with inflation. These results lend strong support to the hypothesis that supply endogenously responds to changes in market size.

Figure 11: Higher Market Size Leads to Product Entry

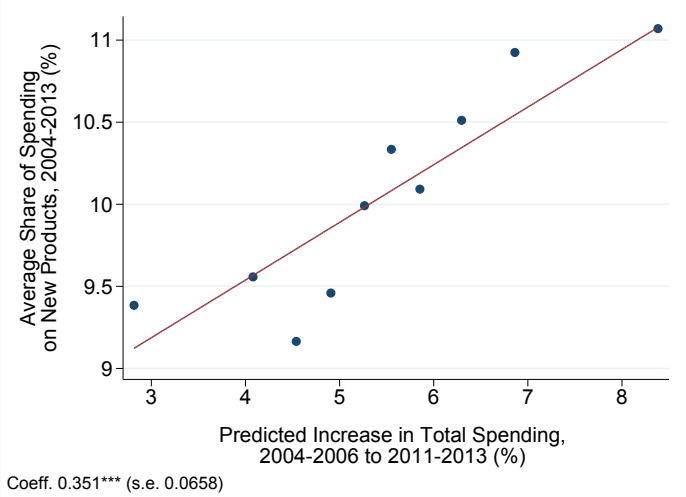


Figure 12: Higher Market Size Leads to Lower Inflation (Overlapping Goods)

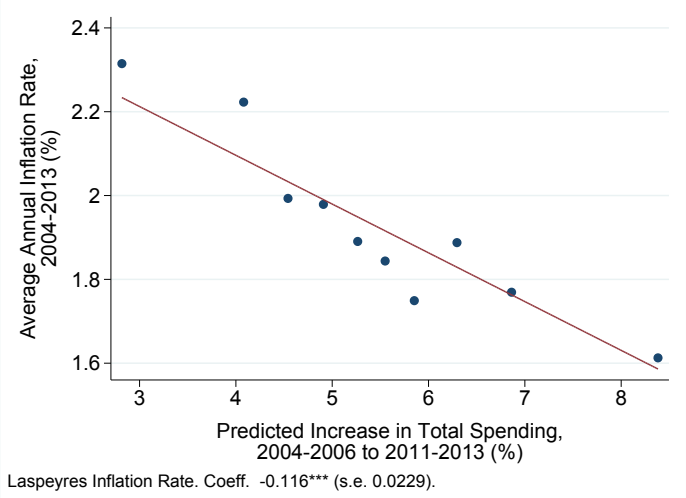


Table 9 shows that the relationships between predicted market size growth, product innovations and inflation are significant at the 1% level. The interpretation of the magnitudes is as follows: a one percentage point increase in the growth of demand[37] causes a 0.35 percentage point increase in the share of spending on new products and a 0.11 percentage point decline in the inflation rate on goods that are available across years. Figure 26 in Appendix D shows the relationship between predicted and actual spending growth, which is also strong. Column 3 of 9 confirms that the relationship between predicted and actual growth of total spending is significant at the 5% level.[38]

[37] where growth of demand is measures as the predicted growth in total spending in a product module - price decile given changes in the age and income distributions.

[38] The point estimate is close to 1, i.e. the predictor is unbiased. Unbiased prediction wasn't necessarily expected, because the measure of actualy total spending growth takes into account both price and quantity effects, while the predicted increase in

Table 9: Causal Effects of Changes in Market Size

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)	Actual Spending Growth (%)
Predicted Increase in Spending (%)	0.351***	-0.116***	1.031**
	(0.0658)	(0.0229)	(0.492)
Product Module Fixed Effects	Yes	Yes	Yes
Spending Weights	Yes	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes	Yes
Number of Observations	9,089	9,089	9,089
Number of Clusters	1,006	1,006	1,006

Standard errors clustered by product modules.

The point estimates are precisely estimated and can be used for a calibration. From the public use micro data from the US Census, I find that between 2004 and 2013 on average the number of high-income households grew 3.12pp faster than the number of low-income households. Multiplying this number by the point estimate in the second column of Table 9 implies an annual inflation difference of 34.3 basis points, which represents 84% of module-decile benchmark[39] and 52% of the overall inflation difference between high- and low-income households. In other words, the response of inflation to changes in market size is sufficiently large to explain most of the difference in inflation rates across the income distribution.

Table 10 shows the robustness of these results. Panel A runs a falsification test in the set of product modules - price deciles that experienced negative spending growth during the period 2004-2013. The model does not predict a significant relationship between change in market size and entry of new products or inflation in this subsample, and indeed I do not find any. Panel B addresses the potential concern that some of the relationship between predicted demand and innovation and inflation could be spuriously driven by a differential increase in supply across the product space. It shows that the results are very similar when considering product module - price deciles that cater to consumers above the age of fifty. In other words, the result is not driven by young consumers, for whom direct supply effects are more likely to exist.

spending is based on the assumption that spending per capita is fixed.

[39]The regressions are all at the product module × price decile level, therefore this is the relevant benchmark.

Table 10: Robustness of Causal Effects of Changes in Market Size

Panel A: Falsification Test in Product Module - Deciles with Negative Spending Growth

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)
Predicted Increase in Spending (%)	-1.093	0.162
	(1.148)	(0.108)
Product Module Fixed Effects	Yes	Yes
Spending Weights	Yes	Yes
Sample Restricted to Negative Spending Growth	Yes	Yes
Number of Observations	632	632
Number of Clusters	305	305

Standard errors clustered by product modules.

Panel B: The Effect Is Not Driven by Young Consumers

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)
Predicted Increase in Spending (%)	0.306***	-0.113***
	(0.075)	(0.021)
Product Module Fixed Effects	Yes	Yes
Spending Weights	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes
Number of Observations	6,571	6,571
Number of Clusters	926	926

Sample restricted to product modules - price deciles with mean consumer age above 50.
Standard errors clustered by product modules.

A variety of additional robustness checks are reported in Appendix D. Panel A shows that the points estimates are very stable when including flexible controls for the initial (2004-2006) age and income distributions in each product module - price decile. Specifically, I control linearly for the 10th, 25th, 50th, 75th and 90th percentiles, as well as the mean, of both the age and income distributions. Panel B also shows stability of the point estimates when introducing fixed effect for each price decile within a product module and when omitting product module fixed effects. Panel B of Table 21 shows that the results are similar when using truncated weights. Statistical significance at the 1% level is maintained in all specifications, with standard errors clustered by product modules.

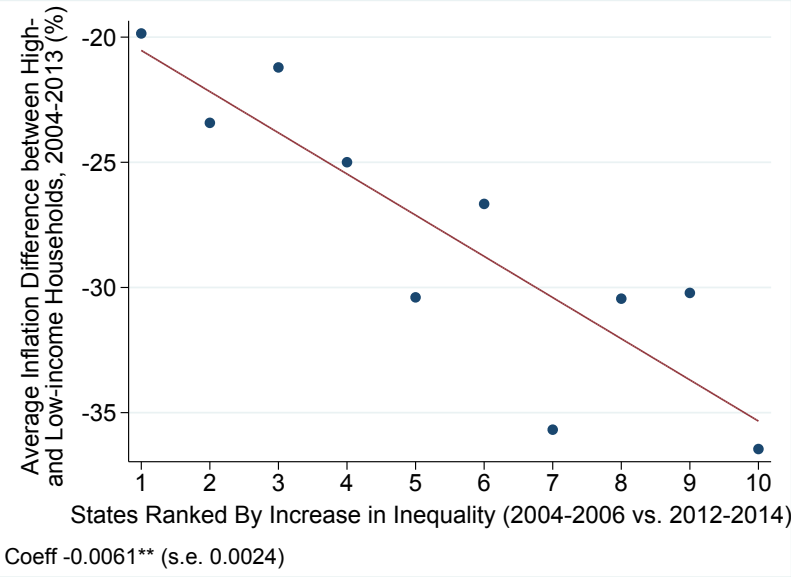
4.4.4 Additional Evidence from Variation Across US States and Local Markets

I provide additional evidence for the relevance of the "market size" hypothesis by exploiting variation in the rate of inequality growth across US states. Using Census public use microdata between 2004-2006 and 2012-2014, I measure the change in the total income accruing to households who earned more than 100k and

less than 30k in each state. Inequality has increased in all 50 states but the rate of increase varied across sates. The increase in inequality was fastest increase in California, Texas and New York and slowest in West Virginia, New Mexico and North Dakota.

UPCs can be thought of as are partly non-tradable because of strength of local brand preferences (Bronnenberg, Dube and Gentzkow, 2012). Therefore, the Nielsen data can be aggregated at the state level to examine how variation in the rate of inequality growth relates to patterns of inflation. In all states, inflation was lower for high-income households earning above $100,000 a year, relative to low-income households making below $30,000 a year. But this difference in inflation rates was relatively larger in states with a faster increase in inequality. Figure 13 shows this result.

Figure 13: The Inflation Difference Between High- and Low-Income Increases as Inequality Increases Faster



In Appendix D, as a robustness check I use time variation in the household age and income distributions in seventy-six local markets tracked by Nielsen within the US between 2004 and 2013. I compare inflation patterns across product module - local market cells with increasing or decreasing predicted market size. I again find that inflation is lower when predicted demand increases - the point estimates are very similar to those obtained from the analysis at the national level and robust to the inclusion of various fixed effects.

4.5 Alternative Mechanisms

I have investigated various alternative explanations for the evidence. First, I show that broad shocks cannot explain the results (shock at the level of MSAs, product groups; recession; rise of online retail; oil shocks). Second, I examine the hypothesis that the production function features increasing returns to scale, using

retailer marginal cost data for a subset of the products in 2004-2007. Third, I study whether dynamic pricing is consistent with the evidence. Finally, I investigate whether trade patterns can explain the results.

Aggregate shocks. First, the various decompositions reported in Section 3 show that the results are not driven by broad shocks that would be specific to certain areas (Appendix Table 20) or to certain departments, product groups or product modules (Tables 1 and 2).

Online retail. The rise of online retail could have differentially benefited high- and low-income households. For instance, if higher-income households are more technology savvy, they might be more likely to use online platforms to search for products, which would increase their price elasticity and result in lower equilibrium markups. However, the inflation difference across product categories is not related to heterogeneity in exposure to online retail - in particular, it persists in categories that were very little affected by online retail during this period, such as food (Table 1).

Innovation dynamics independent of changes in market size. An alternative view of the innovation patterns is that product innovation may always be skewed towards the higher-income consumers, regardless of the underlying patterns of growing inequality. In other words, the patterns documented in Section 3 may be a steady state By introducing flexible controls for the income distribution of consumers and for the quality distribution (price deciles) within a product module, Panel B of Appendix Table 21 shows that the estimated response of product innovations ot market size is not confounded by static patterns related to income or quality. Moreover, I have not found empirical support for the predictions of a simple class of models that generate a steady-state difference in the inflation rates experienced by high- and low-income households - in these models, the equilibrium price elasticity of higher-income consumers should always be lower.[40]

Manufacturer competition. Using the manufacturer identifiers provided by GS1, I have studied patterns of market concentration at the manufacturer level and how they have changed over time. Concentration (as measured by Herfindahl indices) tends to be lower in parts of the product space catering to higher-income households, but changes in concentration over time are small. Overall, the data shows that entry of new manufacturers is not the driving force of the results. This justifies focusing on a model where changes in markups are induced by changes in the elasticity of market demand, rather than by competition between oligopolists, e.g. in a Cournot model.

Household search behavior. Another possible channel for the results is that high-income consumers could have become more price elastic because their search behavior has changed. Such a channel would manifest itself primarily through within-UPC inflation difference between high- and low-income households, which Table 2 shows is not the case.

Other mechanisms. In ongoing work, I use Nielsen TDLink data to characterize changes in the competitive environment of retailers and I document how competitive dynamics differ across areas depending on the density of high- and low- income households. I also examine whether trade is an important channel

[40] Intuitively, if high-income consumers are less price elastic and if the cost of increasing product variety is linear, in equilibrium we will observe a high flow of new products targeting higher income consumers. The equilibrium mechanism is that the high-end products have higher margins (because the high-income consumers are less price elastic) but have a shorter lifecycle (because they get displaced by other high-end product innovations).

by studying heterogeneity in the inflation difference across product groups with different degrees of import penetration. Finally, I test the predictions of models featuring dynamic pricing and increasing returns to scale.

5 Conclusion

In this paper, I have shown that quality-adjusted inflation substantially varies across the income distribution in the retail sector. The current methodology of statistical agencies like BLS cannot capture this variation, which exists primarily at the product level rather than across broad item categories. Furthermore, I have established that product introductions and prices endogenously respond to changes in market size in a way that magnifies the welfare effects of changes in nominal inequality. As shown in a simple calibration, the endogenous response of supply to changes in market size over the past decade can explain most of the observed difference in inflation rates across the income distribution during this period. These findings open up several directions for future research. Do similar results hold beyond the retail sector? How should one adjust optimal redistributive taxation formulas (e.g. as in Mirrlees, 1971) to take into account the endogenous response of supply to changes in market size? These and other extensions await further research.

BIBLIOGRAPHY

Acemoglu, Daron, 2002, "Directed Technical Change," The Review of Economic Studies, Vol. 69, No. 4.

Acemoglu, Daron, 2007, "Equilibrium Bias of Technology," Econometrica, Econometric Society, vol. 75(5), pages 1371-1409.

Argente, David and Munseob Lee, 2015, "Cost of Living Inequality During the Great Recession," Working Paper.

Aguiar, Mark and Erik Hurst, "Consumption versus Expenditure," Journal of Political Economy, 2005, 113 (5), 919–948.

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

Beraja, Martin, Erik Hurst, and Juan Ospina, "The Aggregate Implications of Regional Business Cycles," Working Paper, 2015.

Broda, Christian and David E Weinstein, "Globalization and the Gains From Variety," Quarterly Journal of Economics, 2006, 121 (2), 541–585.

Broda, Christian and David E Weinstein, "Product Creation and Destruction: Evidence and Price Implications," American Economic Review, 2010, 100 (3), 691–723.

Broda, Christian and John Romalis, "TheWelfare Implications of Rising Price Dispersion," 2009.

Chevalier, Judith A and Anil K Kashyap, "Best Prices: Price Discrimination and Consumer Substitution," NBER-Working Paper No. 20768, 2014.

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

Diewert, W Erwin, "Exact and Superlative Index Numbers," Journal of Econometrics, 1976, 4 (2), 115–145.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," American Economic Review, 1994, 84 (1), 157–177.

Gopinath, Gita, Pierre-Olivier Gourinchas, Chang-Tai Hsieh, and Nicholas Li, "International Prices, Costs and Mark-Up Differences." American Economic Review, 2011, 101 (6): 2450–86.

Handbury, Jessie, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across US Cities," 2013.

Hobijn, Bart and David Lagakos, "Inflation Inequality in the United States," Review of Income and Wealth, 2005, 51 (4), 581–606.

McGranahan, Leslie and Anna Paulson, "Constructing the Chicago Fed Income Based Economic Index-Consumer Price Index: Inflation Experiences by Demographic Gropu: 19832-2005," Federal Reserve Bank of Chicago Working Paper.

Pisano, Luigi and Andrea Stella, "Price Heterogeneity and Consumption Inequality," Working Paper.

Sato, Kazuo, "The Ideal Log-Change Index Number," Review of Economics and Statistics, 1976, 58 (2), 223–228.

Stroebel, Johannes and Joseph Vavra, "House Prices, Local Demand, and Retail Prices," NBERWorking Paper No. 20710, 2015.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti and Jacques-Francois Thisse, 2012, "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," Econometrica, Vol. 80, No. 6.

Appendix A

More on Theory

A. The Strong Equilibrium Bias in a Model Following Acemoglu (2002)

Consumers

Individual Demands

Preferences are different for the two groups:

$$U_{poor} = a + u(x_L) = a + \frac{1}{1-\alpha_L} x_L^{1-\alpha_L}$$

$$U_{rich} = a + v(x_R) = a + \frac{1}{1-\alpha_R} x_R^{1-\alpha_R}$$

where a is the numeraire. The FOC for the composite high-quality and low-quality goods are:

$$u'(x_L) = x_R^{-\alpha_L} = p_L$$

$$v'(x_R) = x_R^{-\alpha_R} = p_R$$

Nominal inequality is measured by w_{rich}/w_{poor}, while real inequality is measured by $\frac{w_{rich}}{p_H}/\frac{w_{poor}}{p_L}$.

Aggregate Demands

Normalize the total mass of consumers to 1, with share λ of rich types (if desired, we can introduce another scaling factor to study market size effect due to the total number of consumers). Aggregate demands are given by:

$$D_L = (1-\lambda) p_L^{-\frac{1}{\alpha_L}}$$

$$D_H = \lambda p_H^{-\frac{1}{\alpha_H}}$$

Producers

Final Producer

Final producer just uses capital and combines "varieties $x(v,i)$" to produce two types of goods, H or L (profits are thrown away). With $i = L/H$, we can write the problem as:

$$max_{|x(v,i)| \in N(i)} \frac{p(i)}{1-\epsilon_i} \left(\int_0^{N(i)} x(v,i)^{1-\epsilon_i} dv \right) - \int_0^{N(i)} p^x(v,i) x(v,i) dv$$

Note that the returns to scale are decreasing. The optimal choice is:

$$x(v,i) = \left(\frac{p(i)}{p^x(v,i)} \right)^{\frac{1}{\epsilon_i}}$$

We denote by $\sigma_i = \frac{1}{\epsilon_i}$ the elasticity of substitution between machines.

Intermediate Producers

Intermediate monopolist has a patent and chooses optimal price (we consider one period only here, but easy to extend since problem is separable). The cost of production of a machine is ψ_i units of the final good. The value of a patent (of a variety) for intermediate good in sector i (lasting one period) is:

$$V(v,i) = (p^x(v,i) - \psi_i)x(v,t)$$

where the optimal price chosen by the monopolist maximizes (taking demand as given):

$$max_{p^x(v,i)} \ (p^x(v,i) - \psi_i) \left(\frac{p(i)}{p^x(v,i)}\right)^{\frac{1}{\epsilon_i}}$$

Hence the optimal choice:

$$p^x(v,i) = \frac{\psi_i}{1-\epsilon_i}$$

The value function at the optimum is:

$$V(v,i) = p(i)^{\frac{1}{\epsilon_i}} (\frac{1-\epsilon_i}{\psi_i})^{\frac{1-\epsilon_i}{\epsilon_i}} \epsilon_i$$

Aggregate Supply

Total quantity supplied in equilibrium is given by:

$$S_i^* = \frac{1}{1-\epsilon_i} \left(\int_0^{N(i)} x(v,i)^{1-\epsilon_i} dv \right) = \frac{1}{1-\epsilon} \left(\int_0^{N(i)} \left(\frac{p(i)}{p^x(v,i)} \right)^{\frac{1-\epsilon_i}{\epsilon_i}} dv \right)$$

$$= \left(\frac{1}{1-\epsilon_i} \right)^{\frac{1}{\epsilon_i}} \left(\frac{p(i)}{\psi_i} \right)^{\frac{1-\epsilon_i}{\epsilon_i}} N(i)$$

Solving for the equilibrium

With exogenous varieties

$$p_H = \left[\left(\frac{\lambda}{N_H} \right) (1-\epsilon_H)^{\frac{1}{\epsilon_H}} (\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}} \right]^{\frac{1}{\frac{1-\epsilon_H}{\epsilon_H} + \frac{1}{\alpha_H}}}$$

$$p_L = \left[\left(\frac{1-\lambda}{N_L} \right) (1-\epsilon_L)^{\frac{1}{\epsilon_L}} (\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}} \right]^{\frac{1}{\frac{1-\epsilon_L}{\epsilon_L} + \frac{1}{\alpha_L}}}$$

Relative market size *increases* the relative price, because supply is fixed. So real inequality when prices are endogenous is "lower" than real inequality when prices are exogenous.

With endogenous varieties

Interior solution

At an interior solution (there is research on both kinds of goods), the no arbitrage condition between two types of inventions requires:

$$\eta V(v,H) = \eta V(v,L)$$

The equilibrium ratio of varieties and the equilibrium prices are given by:

$$\frac{(N_H)^{\frac{1}{\epsilon_H}}}{(N_L)^{\frac{1}{\epsilon_L}}} = \frac{(\lambda)^{\frac{1}{\epsilon_H}}}{(1-\lambda)^{\frac{1}{\epsilon_L}}} \left(\frac{\kappa_H}{\kappa_L}\right)^{\zeta} \frac{(\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H^2}}}{(\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L^2}}} \frac{(1-\epsilon_H)^{\left(\frac{1}{\epsilon_H}\right)^2}}{(1-\epsilon_L)^{\left(\frac{1}{\epsilon_L}\right)^2}}$$

$$p_H = (\frac{\psi_H}{1-\epsilon_H})^{1-\epsilon_H} \left(\frac{1}{\epsilon_H}\right)^{\epsilon_H}$$

$$p_L = (\frac{\psi_L}{1-\epsilon_L})^{1-\epsilon_L} \left(\frac{1}{\epsilon_L}\right)^{\epsilon_L}$$

$$V^* = 1$$

Corner solutions and the strong equilibrium bias

$$\begin{cases} V(v,H) > V(v,L) \\ N_H = N + \bar{N_H} \;\&\; N_L = \bar{N_L} \end{cases}$$

$$p_L = \left[\left(\frac{1-\lambda}{\bar{N_L}}\right)(1-\epsilon_L)^{\frac{1}{\epsilon_L}} (\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}\right]^{\frac{1}{\frac{1-\epsilon_L}{\epsilon_L}+\frac{1}{\alpha_L}}}$$

Consider a relative demand shock (compared with previous periods, where the steady state relative demand is embodied in the steady state relative number of varieties).

Assume that all research is allocated to the high quality good:

$$V(v,H) \quad > \quad V(v,L)$$

$$(Profitability\ Condition)$$

$$\iff \quad \frac{\left(\frac{\lambda}{N+\bar{N_H}}\right)}{\left(\frac{1-\lambda}{\bar{N_L}}\right)} > \left[\frac{(\frac{1-\epsilon_L}{\psi_L})^{\frac{1-\epsilon_L}{\epsilon_L}}\epsilon_L}{(\frac{1-\epsilon_H}{\psi_H})^{\frac{1-\epsilon_H}{\epsilon_H}}\epsilon_H}\right]^{\zeta} \frac{(1-\epsilon_L)^{\frac{1}{\epsilon_L}}(\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}}{(1-\epsilon_H)^{\frac{1}{\epsilon_H}}(\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}}$$

There is "overshooting" of the relative price (strong equilibrium bias) if the new relative price is smaller than the old one:

$$\frac{p_H}{p_L} < \frac{\bar{p_H}}{\bar{p_L}} \qquad (Price\ Overshooting\ Condition)$$

$$\iff \quad \frac{\frac{\lambda}{N+\bar{N_H}}}{\frac{1-\lambda}{\bar{N_L}}} < \left[\frac{(\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}}{(\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}} \cdot \frac{(1-\epsilon_L)^{\frac{1-\epsilon_L}{\epsilon_L}}}{(1-\epsilon_H)^{\frac{1-\epsilon_H}{\epsilon_H}}} \cdot \frac{\epsilon_L}{\epsilon_H}\right]^{\zeta} \frac{(1-\epsilon_L)^{\frac{1}{\epsilon_L}}(\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}}{(1-\epsilon_H)^{\frac{1}{\epsilon_H}}(\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}}$$

which cannot be satisfied at the same time as the profitability condition.

B. The Strong Equilibrium Bias in a Model Following Melitz and Ottaviano (2008)

- Preferences for λ high-income consumers and $1-\lambda$ low-income consumers

$$U_i = q_{0i}^c + \alpha \int_{\omega \in \Omega_i} q_{\omega i}^c d\omega - \frac{1}{2}\gamma \int_{\omega \in \Omega_i} (q_{\omega i}^c)^2 d\omega - \frac{1}{2}\eta (\int_{\omega \in \Omega_i} q_{\omega i}^c d\omega)^2$$

$$\Longrightarrow \; p_{\omega i} = \alpha - \gamma q_{\omega i}^c - \eta Q_i^c \; ; \; Q_i = \int_{\omega \in \Omega_i} q_{\omega i}^c d\omega$$

- To enter the differentiated sector, a firm must incur a sunk entry cost of f_E units of labor
  - Then the firm's unit labor requirement of cost c is drawn from a cumulative distribution function $G(c)$ with support on $[0, c_M]$
  - The zero-profit cost cutoff (c_{Di}) is a sufficient statistic that determines firm outcomes as a function of their cost draw:

$$\begin{aligned} p_i(c) &= \frac{1}{2}(c_{Di} + c) \quad (prices) \\ \mu_i &= p_i(c) - c = \frac{1}{2}(c_{Di} - c) \quad (markups) \\ r_i(c) &= \frac{L_i}{4\gamma}[(c_{Di})^2 - c^2] \quad (revenues) \\ \pi_i(c) &= \frac{L_i}{4\gamma}[c_{Di} - c^2] \quad (profits) \end{aligned}$$

- Under the assumption that productivity $\frac{1}{c}$ is Pareto distributed with lower bound $\frac{1}{c_M}$ and shape parameter k, the (closed economy) cost cutoff is given by:

$$c_{Di} = (\frac{\gamma \phi}{L_i})^{\frac{1}{k+2}}$$

  - So the cost cutoff falls (meaning the average productivity is higher) when varieties are closer substitutes (lower γ), when there is a better distribution of cost draws (lower c_M), when sunk costs fall (lower f_E) and in bigger markets (higher L_i). These comparative statics induce an increase in the "toughness of competition" in the form of a larger number of varieties consumed (higher N_i) and lower average prices (lower $\bar{p_i}$).

- This implies that the relative price decreases with (relative) market size
  - Intuition: firms are "locked in" and the long-run supply curve is downward sloping because of entry
  - Note that even in the short run relative prices will never mitigate the increase in inequality
    - Intuition: the marginal cost of production is constant: $p = \frac{\gamma}{L} q + \psi$
    - Can generate price effects mitigating increases in inequality by introducing specialized labor

The remainder of Appendix A is available from the author upon request.

Appendix B

More on Data

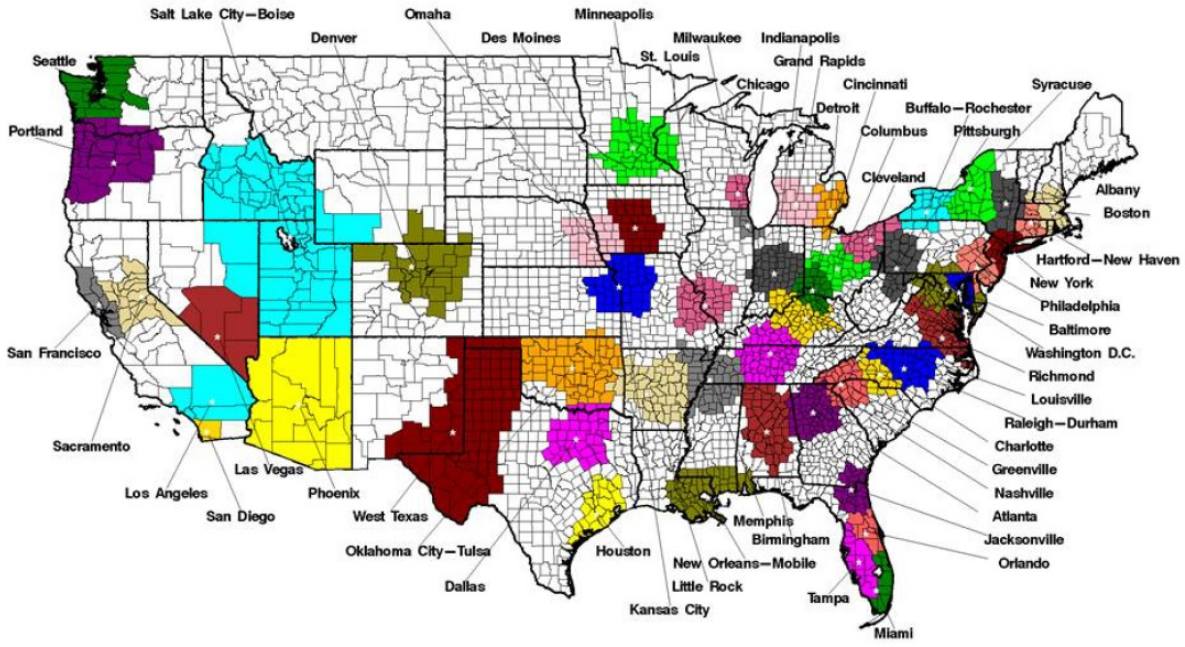
Description of Homescan Consumer Panel Data: I primarily rely on the Home Scanner Database collected by AC Nielsen and made available through the Kilts Center at The University of Chicago Booth School of Business. AC Nielsen collects these data using hand-held scanner devices that households use at home after their shopping in order to scan each individual transaction they have made. Faber and Fally (2015) report that on average each semester covers $105 million worth of retail sales across 58,000 individual, across more than 500,000 barcodes belonging to 180,000 brands.

Description of Retail Scanner Data: The Retail Scanner Data consist of weekly price and quantity information for more than one hundred retail chains across all US markets between January 2006 and December 2013. The database includes about 45,000 individual stores. The stores in the database vary in terms of the channel they represent: e.g. food, drug, mass merchandising, liquor, or convenience stores. Faber and Fally (2015) report that on average each semester covers $110 billion worth of retail sales across 25,000 individual stores, across more than 700,000 barcodes belonging to 170,000 brands.

The strength of the home scanner database is the detailed level of budget share information that it provides alongside household characteristics. Its relative weakness in the comparison to the store-level retail scanner data is that the home scanner samples households and, therefore, has higher sampling error at the product level. Relative to the home scanner data, the store-level retail scanner data records more than one thousand times the retail sales in each semester. I primarily rely on the home scanner data in the paper, but I present robustness checks based on the retail scanner data.

Local Markets: Both the home scanner and retail scanner data can be disaggregated into 76 local markets, which are shown on the map below.

Figure 14: Map of the 76 Local Markets Tracked in the Nielsen Datasets



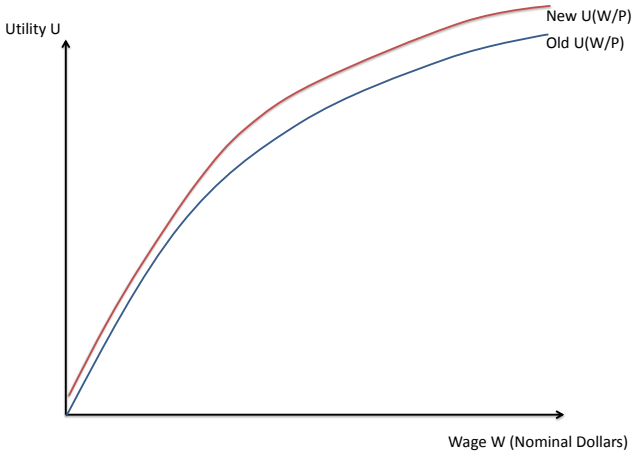
The remainder of Appendix B is available from the author upon request.

Appendix C

Estimation of Quality-Adjusted Inflation and Further Robustness Checks

Nominal and Real Inequality

Figure 15: The Mapping Between Nominal Income and Utility



Estimation Equations

Estimating the elasticities: given the formula reported in the main text, we only need to estimate the group-specific and module-specific elasticities. We do this by first modeling the supply and demand conditions for each good within a module.

The demand equation comes from the following transformation, which exploits the panel nature of the data:

$$\begin{aligned} ln(s_{umgt}) - ln(s_{umg(t-1)}) &= \Delta ln(s_{umgt}) \\ &= (1-\sigma_m)\left[ln(p_{umgt}) - ln(p_{umg(t-1)})\right] + ln(P_{mgt}) - ln(P_{mg(t-1)}) \\ &= (1-\sigma_m)\left[ln(p_{umgt}) - ln(p_{umg(t-1)})\right] + \lambda_{mt} \end{aligned}$$

where the second line uses (1) and the fact that quality/taste is assumed to be constant over time. The fixed effect corresponds to the change in the price index of the module. In practice, there will be an estimation error, which for instance could come from yearly change in taste (which would affect the$d$ parameters). So we can write the demand curve as:

$$\Delta ln(s_{umgt}) = (1-\sigma_m)\Delta ln(p_{umgt}) + \lambda_{mt} + \epsilon_{umgt}$$

Then, we assume an isoelastic supply curve (with $\alpha > 0$ assumed to be the same for all UPCs within a module):

$$ln(c_{umgt}) = \alpha ln(p_{umgt}) + \chi_{mg}$$

$$ln(c_{umgt}) - ln(E_{mgt}) = \alpha ln(p_{umgt}) - ln(E_{mgt}) + \chi_{mg}$$

$$ln(s_{umgt}) = \alpha ln(p_{umgt}) - ln(E_{mgt}) + \chi_{mg}$$

Differencing over time:

$$\begin{aligned} ln(s_{umgt}) - ln(s_{umg(t-1)}) &= \Delta ln(s_{umgt}) \\ &= \alpha \left[ln(p_{umgt}) - ln(p_{umg(t-1)})\right] + ln(E_{mgt}) - ln(E_{mg(t-1)}) \end{aligned}$$

so

$$\begin{aligned} \Delta ln(p_{umgt}) &= \frac{1}{\alpha}\Delta ln(s_{umgt}) - \frac{1}{\alpha}\Delta ln(E_{mgt}) \\ &= \frac{1}{\alpha}\Delta ln(s_{umgt}) + \psi_{mgt} \end{aligned}$$

The fixed effect corresponds to the change in total expenditures in the module (which is observed). In practice there will be estimation error, e.g. due to assembly line shocks, so we write:

$$\Delta ln(p_{umgt}) = \frac{1}{\alpha}\Delta ln(s_{umgt}) + \psi_{mgt} + \delta_{umgt}$$

We now want to eliminate the fixed effects in the demand and supply equations. We take a difference relative to the UPC k with the largest market share:

$$\Delta^k ln(s_{umgt}) = (1-\sigma_m)\Delta^k ln(p_{umgt}) + \epsilon^k_{umgt} \tag{3}$$

$$\Delta^k ln(p_{umgt}) = \frac{1}{\alpha}\Delta^k ln(s_{umgt}) + \delta^k_{umgt} \tag{4}$$

with $\Delta^k X = \Delta X_{umgt} - \Delta X_{kmgt}$, $\epsilon^k_{umgt} = \epsilon_{umgt} - \epsilon_{kmgt}$ and $\delta^k_{umgt} = \delta_{umgt} - \delta_{kmgt}$.

Now we can set up the moment condition, based on the assumption that the upc-specific demand and supply shocks are uncorrelated over time, i.e $E_t[\epsilon^k_{umgt}\delta^k_{umgt}] = 0$.

$$v_{umgt} = \epsilon^k_{umgt} \times \delta^k_{umgt}$$

$$G(\beta_m) = E_t(v_{umgt}(\beta_m)) = 0 \;\; \forall u, m \; and \; g$$

This can be written as:

$$\begin{aligned} v_{umgt}(\beta_m) &= \epsilon^k_{umgt} \times \delta^k_{umgt} \\ &= \left(\Delta^k ln(s_{umgt}) - (1-\sigma_m)\Delta^k ln(p_{umgt})\right) \times \left(\Delta^k ln(p_{umgt}) - \frac{1}{\alpha}\Delta^k ln(s_{umgt})\right) \\ &= \Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt}) - (1-\sigma_m)\left(\Delta^k ln(p_{umgt})\right)^2 - \frac{1}{\alpha}\left(\Delta^k ln(s_{umgt})\right)^2 \\ &\quad + \frac{(1-\sigma_m)}{\alpha}\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt}) \\ &= (\sigma_m - 1)\left(\Delta^k ln(p_{umgt})\right)^2 - \frac{1}{\alpha}\left(\Delta^k ln(s_{umgt})\right)^2 + \frac{\alpha + (1-\sigma_m)}{\alpha}\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt}) \end{aligned}$$

The moment condition $E_t[v_{umgt}(\beta_m)] = 0$ means:

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \frac{1}{\alpha(\sigma_m - 1)} E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \frac{\alpha + (1-\sigma_m)}{\alpha(\sigma_m - 1)} E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right] \quad \forall u, m \text{ and } g$$

If you rewrite $\alpha = \frac{1+\omega_m}{\omega_m}$ to match the notation in Broda and Weinstein (2006), this yields:

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \frac{\omega_m}{(1+\omega_m)(\sigma_m - 1)} E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \frac{(1+\omega_m) + (1-\sigma_m)\omega_m}{(1+\omega_m)(\sigma_m - 1)} E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right]$$

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \frac{\omega_m}{(1+\omega_m)(\sigma_m - 1)} E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \frac{1 - \omega_m(\sigma_m - 2)}{(1+\omega_m)(\sigma_m - 1)} E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right]$$

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \theta_1 E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \theta_2 E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right] \tag{5}$$

We then estimate the parameters ω_m and σ_m, under the restriction that $\omega_m > 0$ and $\sigma_m > 1$. To do this, we first just estimate θ_1 and θ_2 by weighted least squares, as in Feenstra (1994). Then we go back to the primitive parameters. If this produces imaginary estimates or estimates of the wrong sign, we perform grid search for the objective function for values of $\sigma_g \in [1.05, 131.5]$ a intervals that are 5 percent apart.

Average Inflation Rate of Various Income Groups According to Various Price Indices

Table 11: Average Annual Inflation Rates Across Three Income Groups

Panel A: Full Sample (Percentage Points)

	Income $<$ \$30k		Income $\in$ [\$30k-\$100k]		Income $>$ \$100k	
	Arithmetic Avg.	Geometric Avg.	Arithmetic Avg.	Geometric Avg.	Arithmetic Avg.	Geometric Avg.
Geometric Laspeyres	1.212	1.204	0.912	0.951	0.561	0.639
Truncated Geometric Laspeyres	1.544	1.536	1.137	1.157	0.862	0.909
Paasche	1.580	1.571	0.985	1.010	0.965	0.979
Truncated Paasche	1.719	1.710	1.182	1.194	1.117	1.126
Tornqvist	1.938	1.929	1.426	1.418	1.296	1.290
Fisher	1.983	1.974	1.425	1.418	1.327	1.320
Marshall-Edgeworth	1.992	1.984	1.440	1.433	1.330	1.323
CES Ideal	2.041	2.032	1.529	1.522	1.387	1.380
Truncated CES Ideal	2.063	2.054	1.541	1.534	1.413	1.406
Walsh	2.076	2.067	1.571	1.563	1.423	1.416
Truncated Laspeyres	2.257	2.502	1.724	1.910	1.554	1.721
Laspeyres	2.387	2.379	1.867	1.860	1.689	1.682
Truncated Geometric Paasche	2.433	2.424	1.742	1.734	1.822	1.815
Geometric Paasche	2.669	2.660	1.942	1.934	2.037	2.031

Panel B: All Years but Great Recession (Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Geometric Laspeyres	0.870	0.642	0.318
Truncated Geometric Laspeyres	1.179	0.876	0.627
Paasche	1.246	0.732	0.768
Truncated Paasche	1.380	0.928	0.919
Tornqvist	1.586	1.144	1.085
Fisher	1.625	1.161	1.111
Marshall-Edgeworth	1.633	1.176	1.116
CES Ideal	1.674	1.254	1.169
Truncated CES Ideal	1.695	1.268	1.192
Walsh	1.707	1.297	1.204
Truncated Laspeyres	1.891	1.448	1.316
Laspeyres	2.006	1.592	1.455
Truncated Geometric Paasche	2.071	1.467	1.623
Geometric Paasche	2.308	1.648	1.858

Table 11: Average Annual Inflation Rates Across Three Income Groups (*Continued*)

Panel C: Years Prior to Great Recession (Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Geometric Laspeyres	1.210	0.808	0.607
Truncated Geometric Laspeyres	1.545	1.060	0.938
Paasche	1.521	0.906	1.161
Truncated Paasche	1.670	1.186	1.276
Tornqvist	1.854	1.303	1.384
Fisher	1.884	1.281	1.428
Marshall-Edgeworth	1.892	1.294	1.414
CES Ideal	1.966	1.452	1.493
Truncated CES Ideal	2.019	1.493	1.532
Walsh	2.001	1.501	1.536
Truncated Laspeyres	2.249	1.658	1.695
Laspeyres	2.249	1.658	1.695
Truncated Geometric Paasche	2.317	1.688	1.971
Geometric Paasche	2.502	1.802	2.167

Panel D: Years After Great Recession (Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Geometric Laspeyres	0.615	0.519	0.101
Truncated Geometric Laspeyres	0.905	0.738	0.394
Paasche	1.03	0.601	0.473
Truncated Paasche	1.16	0.735	0.651
Tornqvist	1.386	1.024	0.860
Fisher	1.430	1.070	0.873
Marshall-Edgeworth	1.439	1.088	0.892
CES Ideal	1.456	1.106	0.926
Truncated CES Ideal	1.452	1.099	0.937
Walsh	1.485	1.143	0.954
Truncated Laspeyres	1.675	1.321	1.088
Laspeyres	1.823	1.542	1.275
Truncated Geometric Paasche	1.886	1.301	1.362
Geometric Paasche	2.162	1.532	1.625

Table 11: Average Annual Inflation Rates Across Three Income Groups (*Continued*)

Panel E: During the Great Recession (Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Geometric Laspeyres	2.408	1.857	1.411
Truncated Geometric Laspeyres	2.821	2.049	1.685
Paasche	2.751	1.872	1.657
Truncated Paasche	2.903	2.069	1.811
Tornqvist	3.168	2.413	2.036
Fisher	3.235	2.351	2.081
Marshall-Edgeworth	3.249	2.364	2.081
CES Ideal	3.323	2.492	2.147
Truncated CES Ideal	3.352	2.498	2.186
Walsh	3.369	2.531	2.189
Truncated Laspeyres	3.721	2.831	2.506
Laspeyres	3.721	2.831	2.506
Truncated Geometric Paasche	3.700	2.705	2.518
Geometric Paasche	3.933	2.973	2.666

Table 12: Average Annual Inflation Rates Across Four Income Groups

Panel A: Full Sample (Percentage Points)

	Income $<$ \$25k	Income $\in$ [\$25k-\$50k]	Income $\in$ [\$50k-\$100k]	Income $>$ \$100k
Geometric Laspeyres	1.236	1.029	0.785	0.561
Truncated Geometric Laspeyres	1.561	1.293	1.025	0.862
Paasche	1.647	1.249	0.962	0.965
Truncated Paasche	1.766	1.414	1.132	1.117
Tornqvist	2.000	1.668	1.365	1.296
Fisher	2.045	1.687	1.377	1.327
Marshall-Edgeworth	2.052	1.698	1.396	1.330
CES Ideal	2.086	1.763	1.462	1.387
Truncated CES Ideal	2.106	1.778	1.474	1.413
Walsh	2.116	1.800	1.501	1.423
Truncated Laspeyres	2.293	1.984	1.657	1.554
Laspeyres	2.445	2.126	1.795	1.689
Truncated Geometric Paasche	2.527	2.090	1.738	1.822
Geometric Paasche	2.769	2.311	1.949	2.037

Panel B: All Years but Great Recession (Percentage Points, Arithmetic Average)

	Income $<$ \$25k	Income $\in$ [\$25k-\$50k]	Income $\in$ [\$50k-\$100k]	Income $>$ \$100k
Geometric Laspeyres	0.843	0.729	0.529	0.318
Truncated Geometric Laspeyres	1.164	1.007	0.769	0.627
Paasche	1.289	0.964	0.723	0.768
Truncated Paasche	1.405	1.148	0.885	0.919
Tornqvist	1.613	1.374	1.097	1.085
Fisher	1.660	1.392	1.127	1.111
Marshall-Edgeworth	1.666	1.403	1.148	1.116
CES Ideal	1.692	1.469	1.201	1.169
Truncated CES Ideal	1.713	1.489	1.211	1.192
Walsh	1.722	1.504	1.241	1.204
Truncated Laspeyres	1.895	1.683	1.394	1.316
Laspeyres	2.033	1.823	1.533	1.455
Truncated Geometric Paasche	2.143	1.805	1.474	1.623
Geometric Paasche	2.388	2.024	1.669	1.858

Table 13: Average Annual Inflation Rates across the Income Distribution at UPC*Geography Level (Full Sample, Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Paasche	2.065	1.401	1.341
CES Ideal	2.434	1.902	1.722
Laspeyres	2.789	2.365	2.08

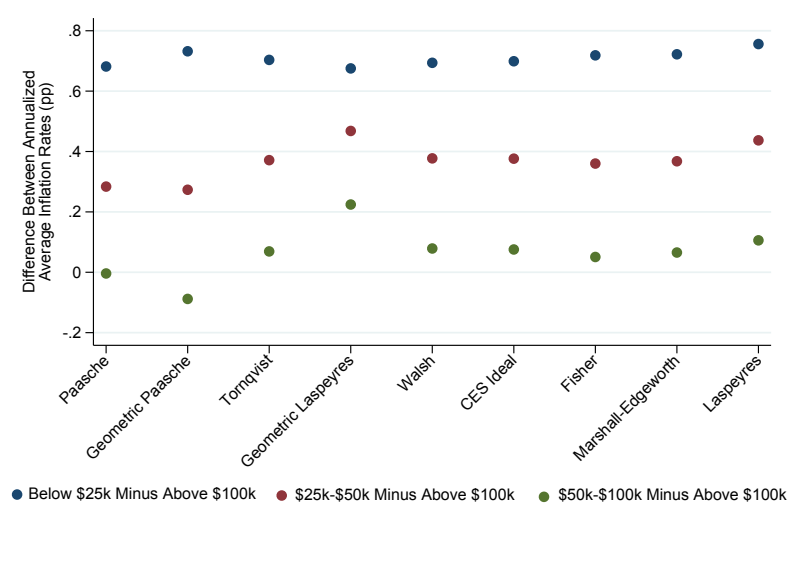
Table 14: Average Annual Inflation Rates across the Income Distribution at UPC*Store Level (Full Sample, Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Paasche	2.239	2.002	1.692
CES Ideal	2.471	2.248	1.901
Laspeyres	2.710	2.471	2.072

Table 15: Average Annual Inflation Rates across the Income Distribution at Quarterly Level (Full Sample, Percentage Points, Arithmetic Average)

	Income $<$ \$30k	Income $\in$ [\$30k-\$100k]	Income $>$ \$100k
Paasche	-1.161	-2.268	-2.124
CES Ideal	1.911	1.107	1.066
Laspeyres	5.429	5.042	4.956

Figure 16: Inflation Difference Between Various Income Groups For Various Price Indices (Fixed Basket)



Is Differential Inflation (Fixed Basket) Across the Income Distribution Driven by a Selection Effect?

Table 16: Products that are about to exit have a lower inflation rate

Subsample	Laspeyres Inflation Rate	Median Laspeyres Inflation Rate (Across Product Modules)
Continued	2.03%	2.06%
About to Exit	-1.33%	-0.52%
Justed Entered	0.03%	1.3%

Table 17: Products that are about to exit have a higher price level

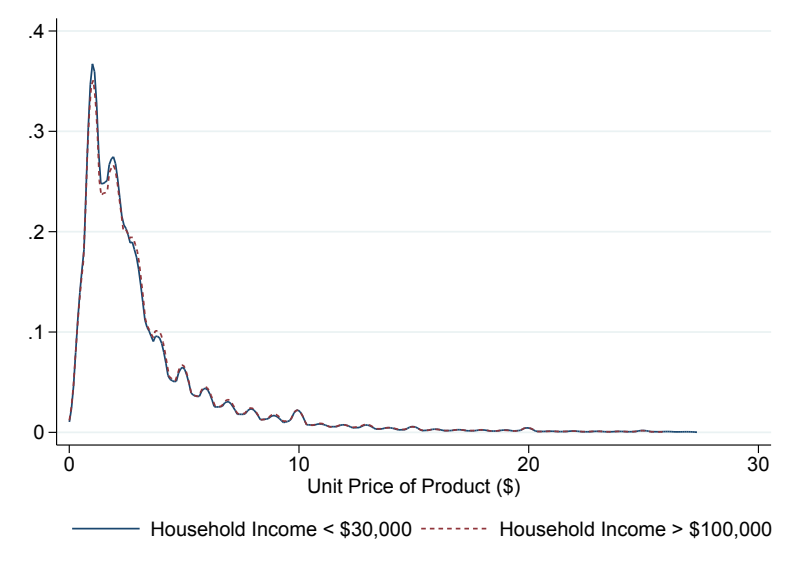
Subsample	Average Price Level	Median Price Level (Across Product Modules)
Continued	3.67	2.75
About to Exit	3.95	2.68
Justed Entered	4.91	3.05

Table 18: Share of spending on new and discontinued products across the income distribution

	Share of Spending on Products...	
Household Income	About to Exit	Just Entered
$> \$100,000$	3.04%	10.94%
$\$30,000 - \$100,000$	2.71%	10.01%
$< \$30,000$	2.59%	9.26%

Differences in Prices Paid for Same Products for Rich and Poor

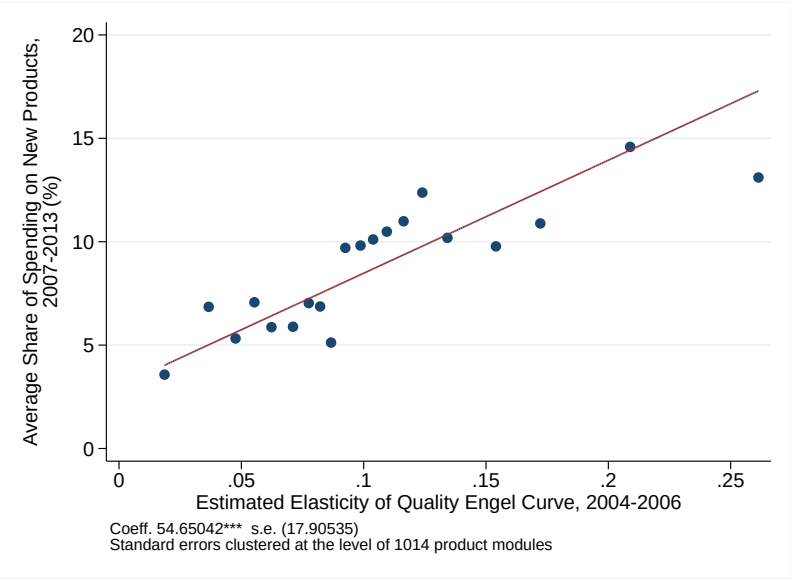
Figure 17: The Distribution of Average Unit Prices Paid is the Same Across the Income Distribution (Reweighting by Spending Shares)



Further Robustness Checks on Increase in Product Variety Across the Income Distribution

Figure 18: The Positive Relationship Between Share of Spending on New Products and Mean Consumer Income

Panel A: Across Product Modules



Panel B: Across Product Modules with Product Group Fixed Effects

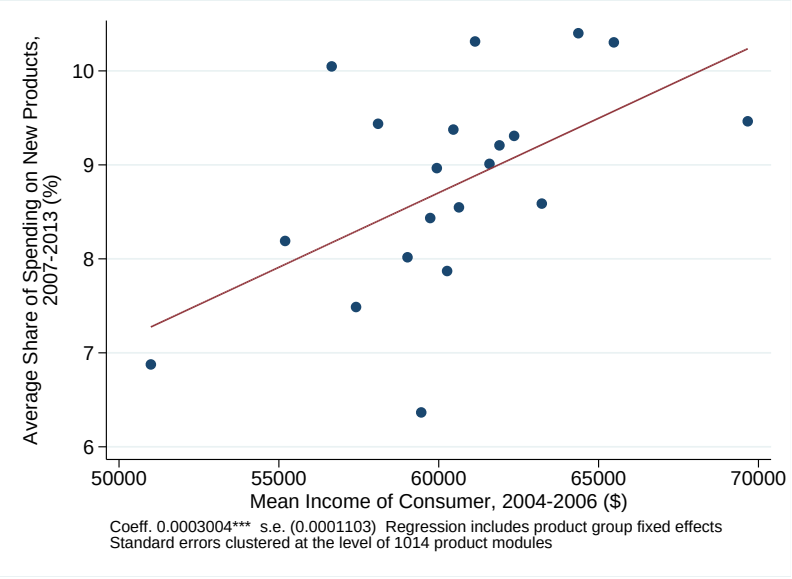
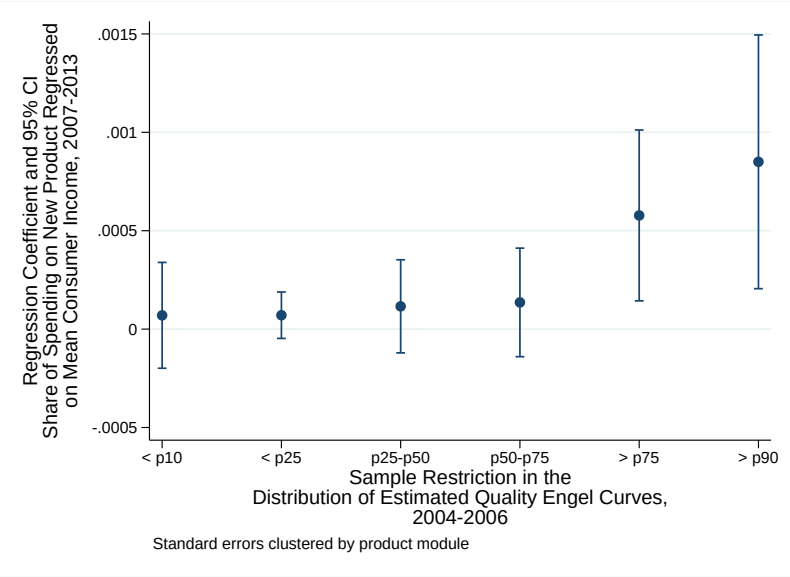


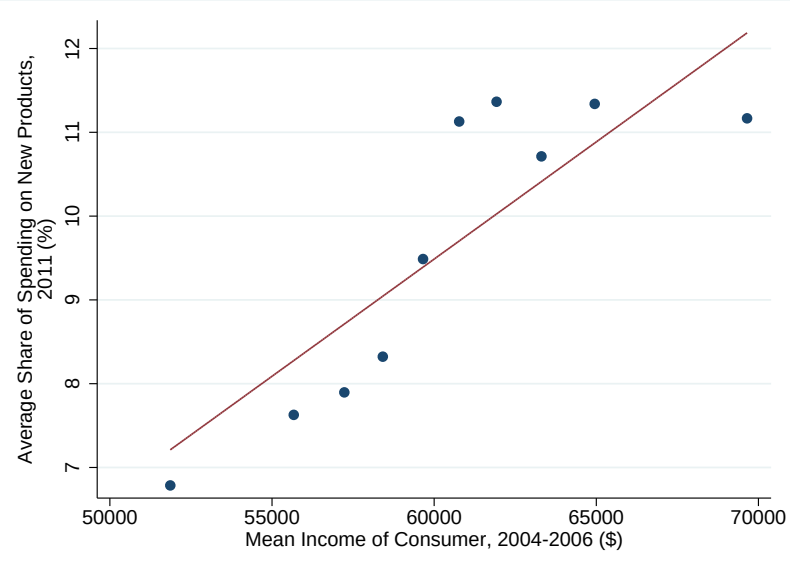
Figure 19: The Relationship Between Share of Spending on New Products and Mean Income Depends on the Quality Engel Curves Elasticity



.

Figure 20: The Relationship Between Share of Spending on New Products and Representative Consumer Mean Income, Controlling for Household Fixed Effects

Panel A: Relationship in the Full Sample



Panel B: Relationship across the Distribution of Quality Engel Curve Elasticities

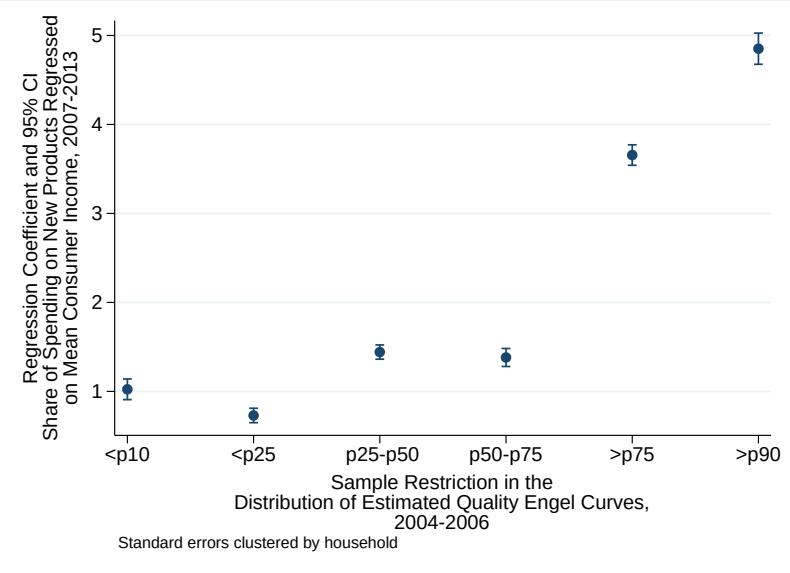
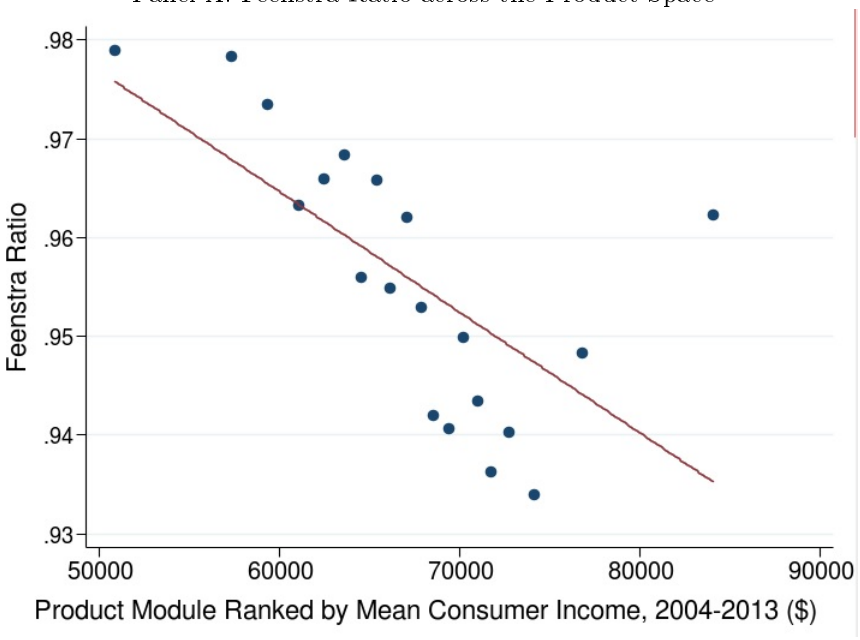


Figure 21: Robustness Checks on Increase in Product Variety across the Income Distribution

Panel A: Feenstra Ratio across the Product Space



Panel B: Annual Growth in Total UPC Count across the Product Space

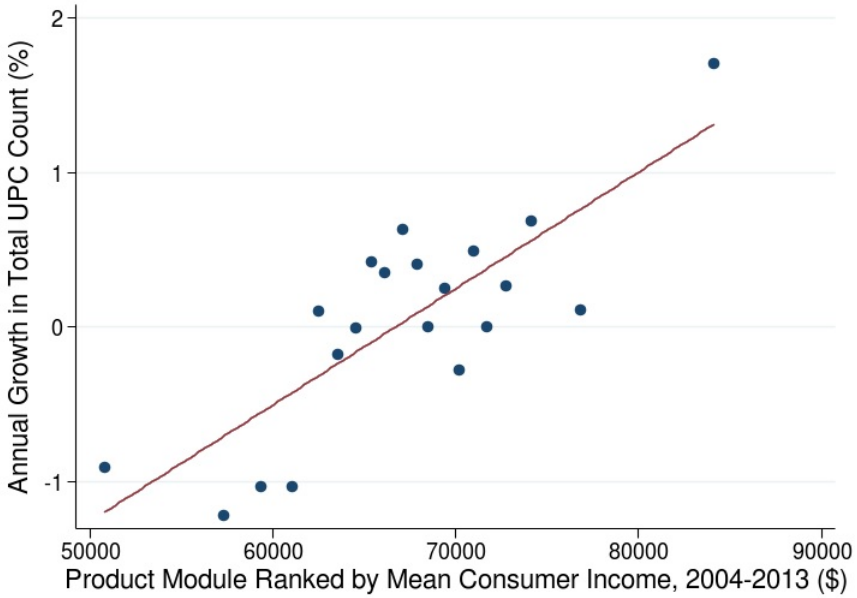


Table 19: Decomposing the Difference in Sahares of Spending between High- and Low-Income Households

Aggregation Level (Broad to Narrow)	Decomposition	Difference (% of actual)
Department	Between	**1.8**
Product Group	Between	**29.0**
Product Module	Between	**39.2**

Quality-Adjusted Inflation

Figure 22: Elasticities of Substitution Differ Across the Income Distribution

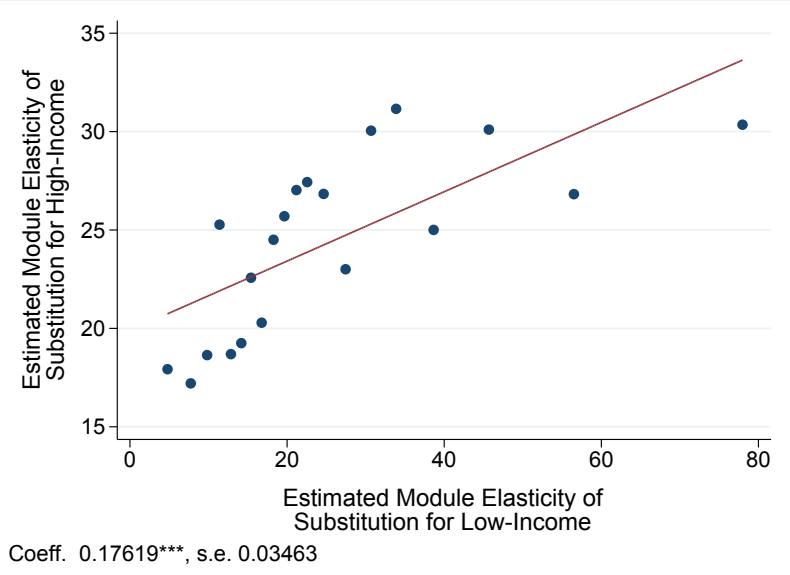
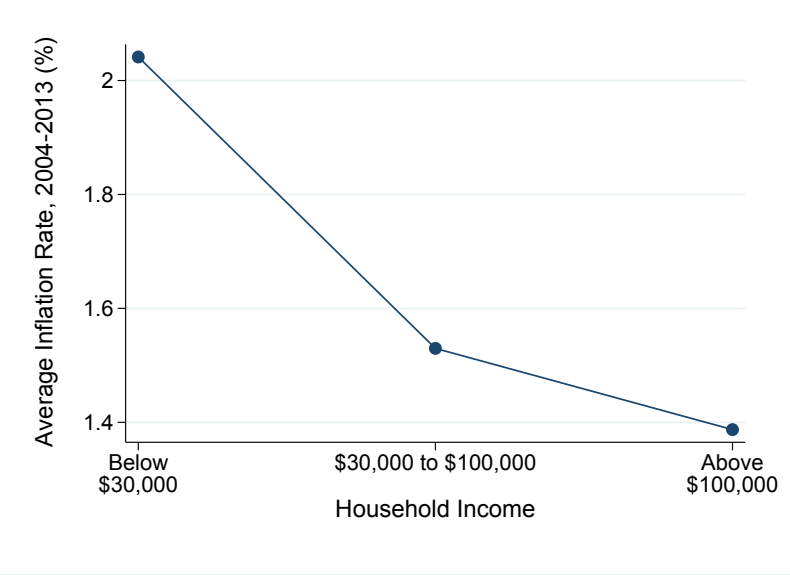


Figure 23: Quality-Adjusted Inflation Across the Income Distribution



Further Robustness Checks

Table 20: Isolating the Contribution of Stores and Local Market to the Overall Inflation Difference

Price Change	Local Market Shares	Store Shares	UPC Shares	Inflation Difference (% of Benchmark)
Counterfactual	Actual	Actual	Actual	**100**
Counterfactual	Counterfactual	Actual	Counterfactual	**43.2**
Counterfactual	Actual	Counterfactual	Counterfactual	**3.1**

Notes: This table is based on a reweighting methodology that expresses the spending share of a product as a combination of the spending shares of teh local market, of the store within that market, and of the barcode within that store. The counterfactal share explained by store effects is an upper bound because for some stores, only high-income or only low-income consumers are observed.

$$P_L \equiv \sum_{i=1}^{n} \frac{p_i^t}{p_i^0} s_{local\ marlet} \cdot s_{store} \cdot s_{upc}$$

The remainder of Appendix C is available from the author upon request.

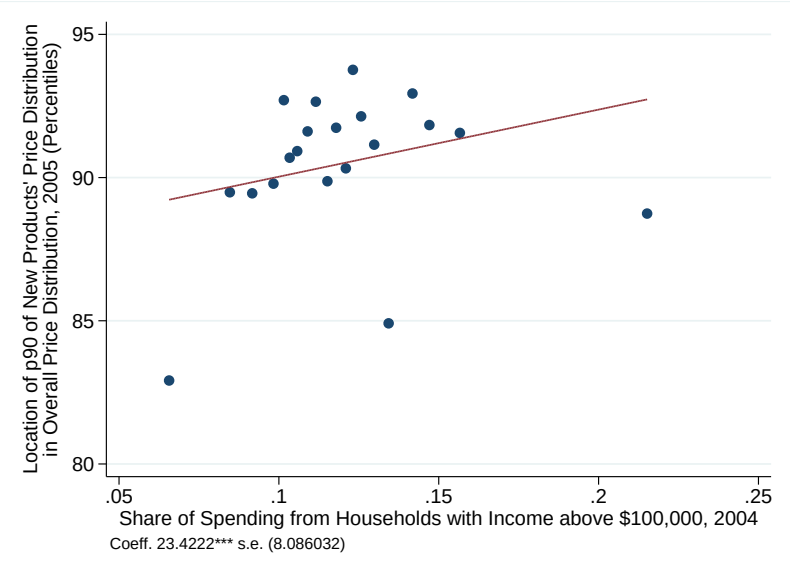
Appendix D

Robustness Checks on The Relationship Between Changes in Market Size and Quality-Adjusted Inflation

Additional Motivating Evidence

Figure 24: The Share of Rich Households is Primarily Correlated with High Price Quantiles of New Products

Panel A: Location of p90 of New Products in Overall Price Distribution



Panel B: Location of p50 of New Products in Overall Price Distribution

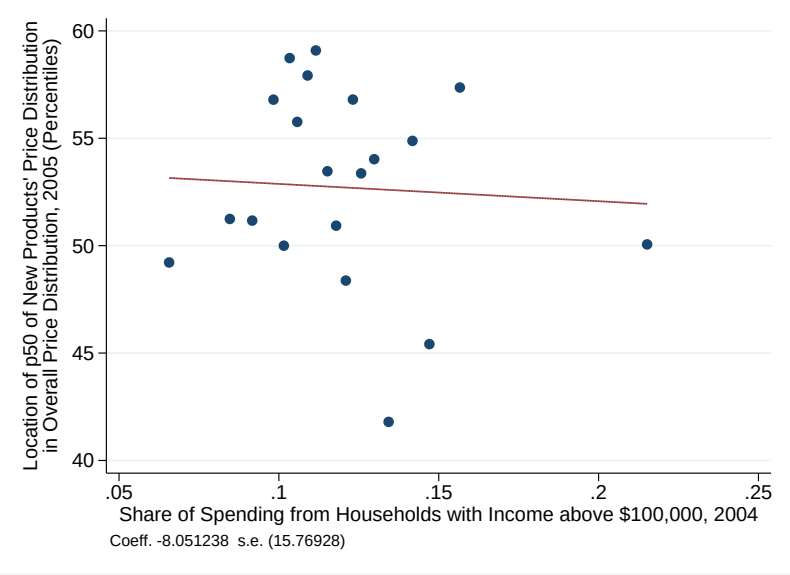
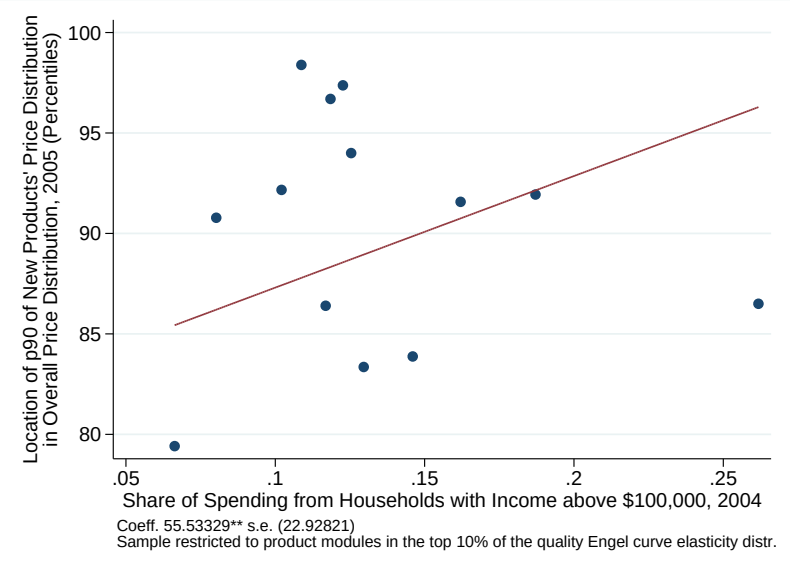
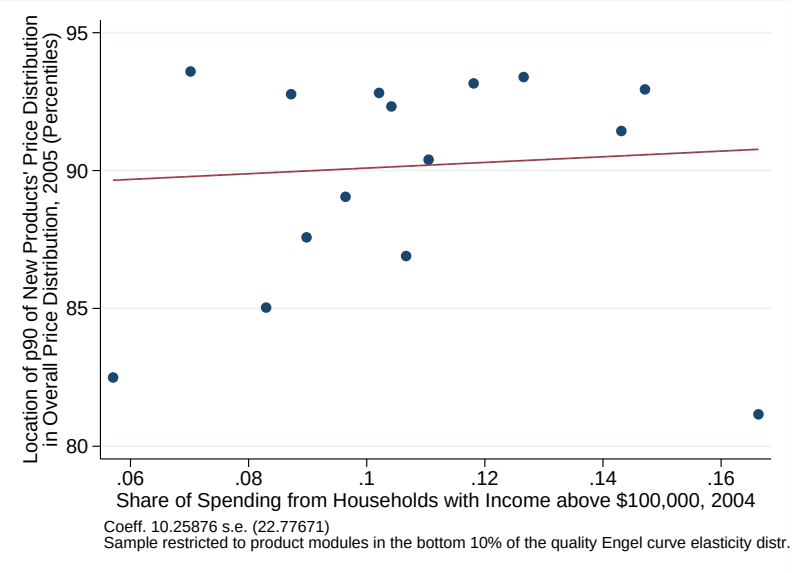


Figure 25: The Share of Rich Households is Correlated with High Price Quantiles of New Products only in Product Modules with Steep Quality Engel Curves

Panel A: Location of p90 of New Products in Overall Price Distribution for Product Modules with Steep Quality Engel Curves



Panel B:Location of p90 of New Products in Overall Price Distribution for Product Modules with Flat Quality Engel Curve



Additional Results on National Research Design

Figure 26: Predicted and Actual Increase in Market Size

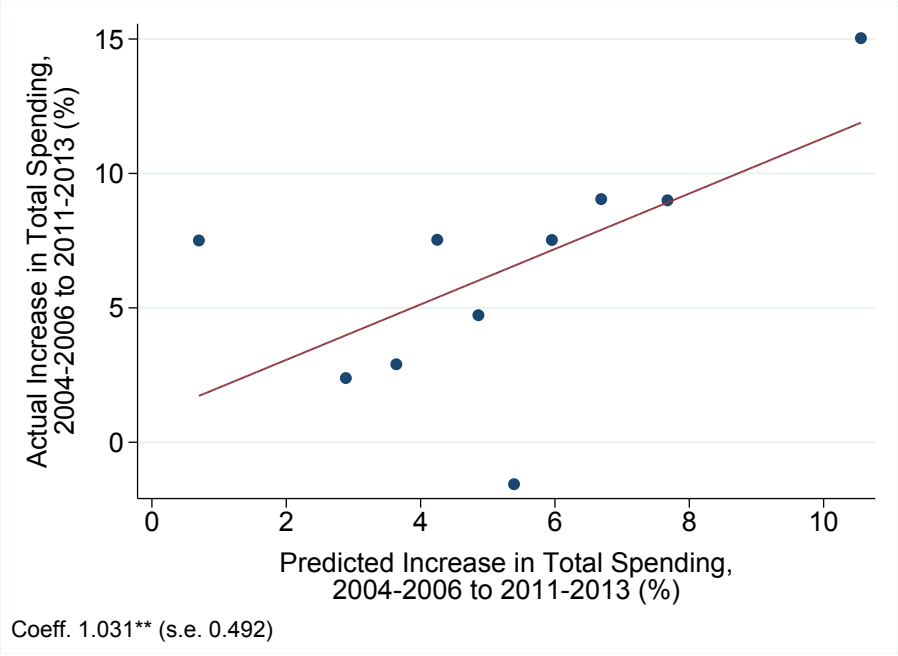


Table 21: Further Robustness Checks on Causal Effects of Changes in Market Size

Panel A: Controlling for 2004-2006 Age and Income Distributions and Price Decile Fixed Effects

	Share of Spending on New Products (pp)			Overlapping Goods Inflation Rate (pp)		
Predicted Increase in Spending (%)	0.527***	0.380***	0.419***	-0.159***	-0.137***	-0.125***
	(0.072)	(0.144)	(0.179)	(0.022)	(0.037)	(0.038)
Age Distribution Controls	Yes	Yes	Yes	Yes	Yes	Yes
Income Distribution Controls	No	Yes	Yes	No	Yes	Yes
Price Decile Fixed Effects	No	No	Yes	No	No	Yes
Product Module Fixed Effects	No	No	No	No	No	No
Spending Weights	Yes	Yes	Yes	Yes	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	9,089	9,089	9,089	9,089	9,089	9,089
Number of Clusters	1,006	1,006	1,006	1,006	1,006	1,006

Standard errors clustered by product modules.

Panel B: Robustness to Other Weights

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)
Predicted Increase in Spending (%)	0.296***	-0.086***
	(0.0585)	(0.0155)
Product Module Fixed Effects	Yes	Yes
Spending Weights	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes
Number of Observations	8,545	8,545
Number of Clusters	1,000	1,000

Sample restricted to product modules below 95th percentile of total spending.
Standard errors clustered by product modules.

Results from Geography Research Design

As a robustness check, I use time variation in the age and income distribution of households in 76 local markets tracked by Nielsen within the US (see Appendix B for details). A local market is a county group defined by Nielsen, which I match to local covariates from the American Community Survey. For each product module in each local market, I predict change in market size based on local change in age and income distributions. Some cities like San Francisco have experienced an increasing share of high-income and young households, while in other cities like New Orleans have become poorer, with a decline in overall population. I then compare the change in fixed basket inflation across product module× local market cells with increasing or decreasing predicted market size. To control for supply factors, I include fixed effects: local market fixed effects control for local scale effects, while product group fixed effect control for national trends in inflation.

In this setting, the identification assumption is that, conditional on the fixed effects, the direct effect of local changes in the age and income distributions on the equilibrium is only through demand.

I use 18 covariates X_{it} (all expressed in logs): total number of households, total population, total female population, total male population, total number of households in age $\times$ income groups (considering four age groups - below 25, 25 to 44, 45 to 64, above 65 -, and dividing each group into 3 income groups - below \$30k, \$30k to \$100k, above \$100k), median household income and mean household income.

Formally, I consider two periods, 2004-2006 and 2011-2013, and I predict (log) local total expenditures Q_{MIT} with local market covariates and fixed effects:

$$Q_{MIT} = \beta^M \cdot 1_M \cdot X_{IT} + \gamma_{IT} + \delta_{GT} + \epsilon_{IMT}$$

where M denotes the product module, G the product group, I the local market, and T the period. Note that the β^M coefficients are allowed to freely vary across modules (i.e. some modules will be very responsive to the number of low-income households, others more responsive to the number of high-income and old households, etc.). I estimate this specification in $2004-2006$ and predict market size out-of-sample in $2011-2013$ (the R^2 is very high: see Appendix D).

The predictor of residual market size growth between two periods is therefore $\beta^m \cdot 1_M \cdot (X_{IT_2} - X_{IT_1})$. Finally, I run specifications of the form:

$$Y_{MI} = \alpha \left[\beta^m \cdot 1_M (X_{IT_2} - X_{IT_1})\right] + \widetilde{\gamma_I} + \widetilde{\delta_G} + \widetilde{\epsilon_{IM}}$$

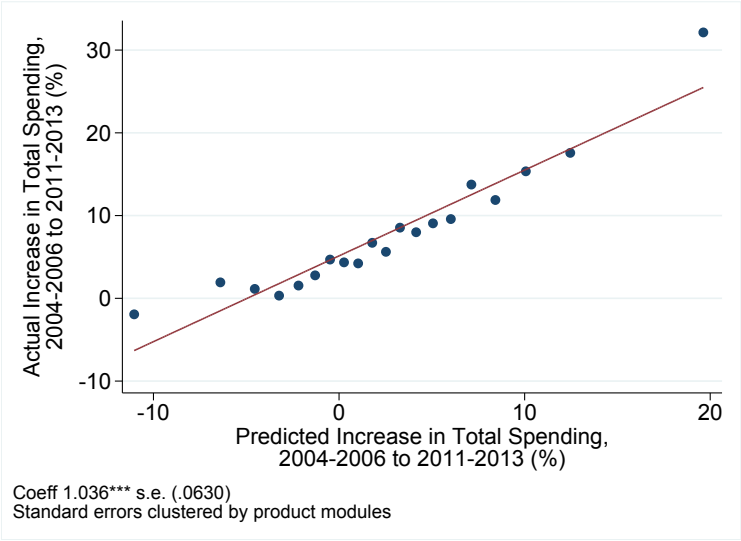
Figure 27 and the table below show that the relationship is very stable and very similar to the results found from the variation at the national level.

Table 22: Causal Effects of Changes in Market Size (Local Level)

	Difference in Fixed Basket Inflation Rate (pp)			
Predicted Increase in Spending (%)	-0.1471***	-0.1276***	-0.1271***	-0.1276***
	(0.0162)	(0.0172)	(0.0188)	(0.0259)
F.E.	Department	Product Group	Product Group	Product Group
Weights	Yes	Yes	No	Yes
Cluster	Local Market	Local Market	Local Market	Product Module

Figure 27: Causal Effects of Changes in Market Size (Local Level)

A. Predicted and Actual Increase in Market Size



B. Higher Market Size Leads to Lower Inflation (Fixed Basket)

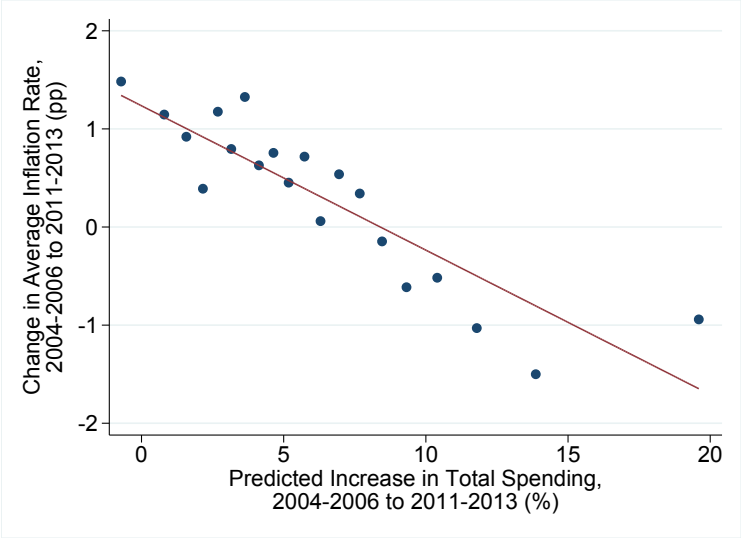
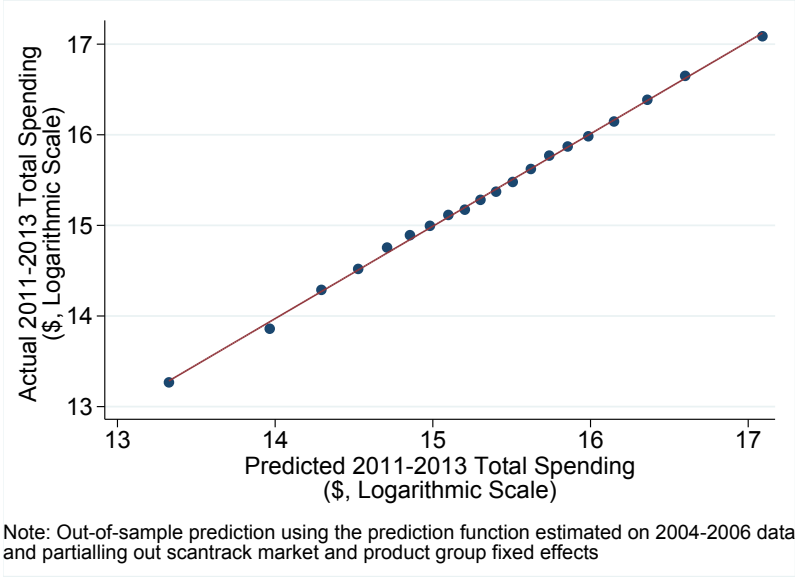
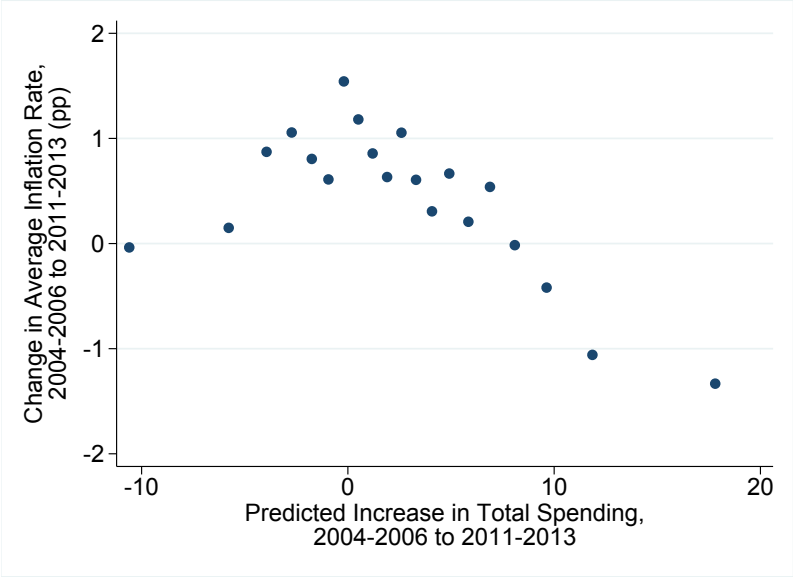


Figure 28: Geography Design: Predicting Market Size



Note: Out-of-sample prediction using the prediction function estimated on 2004-2006 data and partialling out scantrack market and product group fixed effects

Figure 29: Geography Design: Predicted Market Size Growth and Fixed-Basket Inflation



Additional Evidence on Mechanism

Figure 30: Do Product Innovations Follow Market Size or Change in Market Size?

	Share of Spending on New Products	
Lagged Change in Market Size	3.107***	1.901**
	(1.139)	(0.926)
Lagged Market Size	1.399	0.577
	(1.439)	(1.269)
Product Group Fixed Effects	No	Yes
Weights	Yes	Yes

The remainder of Appendix D is available from the author upon request.