Geospatial Heterogeneity in Inflation: A Market Concentration Story*

Seula Kim† — Penn State and IZA

Michael A. Navarrete‡ — NBER

February 17, 2025

[Click for the Latest Version]

Abstract

We study how spatial variation in inflation affects real income inequality and the role of retailer market structure in driving disparities. Using the NielsenIQ Retail Scanner dataset and the Business Dynamic Statistics, we document new stylized facts of spatial heterogeneity in inflation and retailer market structure. We find that poorer MSAs experienced higher food inflation than richer ones, with an annualized gap of 10 p.p. annual gap from 2006 to 2020. Poorer areas also had fewer goods, fewer retailers, and higher market concentration. Using a triple-difference estimator during the 2014-2015 bird flu outbreak, we identify the causal link between retailer concentration and inflation disparities. We build a model with a nested CES structure and Bertrand competition, suggesting that retailer market power is a potential source behind the linkage, and provide data evidence ruling out alternative cost-driven explanations.

JEL Code: E31, I31, L11, L81

Keywords: inflation, spatial inequality, market concentration, retailer market structure

*This paper was previously circulated under the title "Geospatial Heterogeneity in Inflation: The Role of retailer market structure." We thank Miguel Ampudia (discussant), Ben Bernanke, Marcus Casey, Thesia Garner, John Haltiwanger, Judith Hellerstein, Colin Hottman (discussant), David Johnson, Ethan Kaplan, Munseob Lee, Giacomo Ponzetto, Lumi Stevens, and participants at seminars at BLS, University of Maryland, Federal Reserve Board, Federal Reserve Bank of Atlanta, and various conferences including the NBER Competition in the U.S. Agricultural Sector Meeting, BSE Summer Forum, UChicago-SNU Inequality conference, BdE-BOC-ECB-Fed Conference, NASMES Meeting, DC Urban Economics Workshop, and Midwest Macro Meeting for helpful comments. We also thank the RESET team for providing the infrastructure to address our research project. This research is funded by the Washington Center for Equitable Growth. Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Michael Navarrete worked on this paper while a PhD student at the University of Maryland, College Park. All errors are our own.

†Email: seulakim@psu.edu. Address: 614 Kern Building, University Park, PA 16802.

‡Email: navarrem@nber.org. Address: National Bureau of Economic Research Cambridge, MA 02138.

1 Introduction

Inflation is a key economic indicator with significant implications for growth, stability, and the cost of living. However, it is often measured and studied at the aggregate level, particularly within disaggregated food categories. Literature and policymakers tend to overlook the heterogeneity in inflation rates across regions, which may mask important regional disparities.

Understanding local variations in food inflation is important for various reasons. Households in different regions experience varying price changes and adjust their consumption patterns based on local prices.[1] In particular, food markets are segmented and localized compared to other products, such as consumer technology, where local market structures can have a significant impact on price variation.[2] Also, food is a necessity and constitutes a substantial portion of household budgets, especially for lower-income and vulnerable households that spend a larger share of their income on food.[3] Therefore, spatial variations in food inflation gives important implications for consumer welfare and spatial inequality, as well as for policymakers to design more effective place-based policies. However, this has been underexplored in literature.

In this paper, we aim to address this gap by documenting spatial heterogeneity in food inflation rates and exploring the role of retailer market structure and its aggregate implications. We use NielsenIQ Retail Scanner data, which provides granular details of 12-digit universal product codes (UPCs), to construct price indexes for disaggregated personal consumption expenditure (PCE) food items at the metropolitan statistical area (MSA) level.[4]

In the data, we reveal several new stylized facts. First, food inflation rates vary across regions with different income levels. Poor MSAs show higher inflation rates than wealthier MSAs on average from 2006 to 2020, with a cumulative difference of about 10 percentage points between the bottom and top deciles. This pattern holds for both disaggregated and aggregated food items, and even when restricting the sample to common UPCs sold across all 10 deciles, referred to as the

[1] This is partly because moving is costly with migration rates declining since the 1980s (Kristin Kerns-D'Amore and McKenzie, 2022).

[2] In the Nielsen Consumer Panel, we find that 92% of households purchase food items exclusively within their residential MSAs. More details are provided in Appendix A.

[3] Schanzenbach et al. (2016) find in the Consumer Expenditure Survey that low-income households allocate a larger share of their budget to food than middle- or high-income households. Specifically, low-income households spend nearly 20% of their expenditures on food, compared to 13% for middle-income households and an even smaller share for high-income households.

[4] For example, two cans of Campbell's tomato soup in different sizes would be classified as two different UPCs.

"common goods rule." This restriction helps eliminate the effect of varying consumption baskets across MSAs with different income levels.

Second, we find that product and store varieties vary across regions. Richer areas have more varieties of goods (UPCs) and a greater number of stores and chains. Typically, the UPCs available in poorer deciles are a subset of those in wealthier deciles. This suggests that imposing the common goods rule across MSA deciles limits the UPCs in wealthier areas but has minimal impact on the basket of goods available in poorer areas.

Third, we show that regions with different income levels exhibit heterogeneous retailer market structures. In the NielsenIQ Retail Scanner data, we define large and small retailers by the top and bottom deciles based on the number of stores nationally. We find that in poorer areas, the share of large retailers is higher, while the share of small retailers is smaller. The opposite pattern is observed in richer areas. Alternatively, we use the Business Dynamic Statistics (BDS) and define large and small retailers with employment size.[5] The results are robust with this alternative definition. Additionally, retailer sales are lower and more concentrated in poorer areas.

Next, we investigate the relationship between inflation and retailer market structure. Using OLS regression, we find that market concentration is associated with higher inflation rates at the MSA level. However, this does not establish causal relationship. In order to identify causality, we use a quasi-experiment of the 2014-2015 bird flu episode as an exogenous supply shock to study its effect on egg price inflation. In particular, we apply a triple difference-in-difference estimation and examine the effect on inflation in MSAs impacted by the outbreak with higher market concentration. We find that the treated MSAs with higher market concentration (measured by sales HHI) experience higher inflation rates than those with lower market concentration. This identifies the causal impact of higher market concentration on regional inflation rates. Our findings suggest that concentrated retail markets exacerbate regional disparities in inflation, particularly during supply shocks.

Lastly, we investigate potential mechanisms behind the causal link between retailer market concentration and inflation rates. One hypothesis is through retailer market power. We build a simple model with a nested CES structure and Bertrand competition, where retailer markups depend on market share. The model suggests that retailer markups are higher in regions with higher market concentration, which contributes to higher inflation rates in these areas. Alternatively, inflation

[5] Large retailers are those with 500 or more employees, while small retailers have 19 or fewer employees.

disparities could stem from cost differentials across regions. If retailer marginal costs grow faster in poorer areas, this could also drive inflation. However, we find suggestive evidence from local wage growth patterns that can rule out this cost-driven explanation.

These findings have important implications for policymakers in multiple dimensions. First, the regional variation in inflation suggests that real income inequality, when assuming uniform inflation across the U.S, underestimates the actual disparity. Specifically, the real income gap between the top and bottom deciles would widen if regional price deflators are used, with 2006 real income as the baseline. This underscores the need for policymakers to adopt localized approaches to measuring inflation and assess regional disparities when designing policies to address income inequality more accurately and effectively.

In addition, comparing our index to the official PCE price index from the Bureau of Economic Analysis (BEA), we find that the official index closely mirrors our index for the top MSA decile. This is because the official indexes are aggregated at the national level using expenditure weights, which are disproportionately accounted for by rich areas. As a result, relying on aggregate indexes may underrepresent inflation experienced in poorer regions while overrepresenting inflation in wealthier areas. This can potentially misinform inflation measurements in low-income regions and lead to less accurate policy assessments.

Furthermore, our spatial focus highlights food market segmentation and retailer market structure as important sources of policy implications. Given the localized nature of food markets, with a large share of purchases made locally, higher food inflation can directly impact the welfare of consumers in local markets. In particular, it can disproportionately affect vulnerable and immobile individuals in low-income areas, who spend a higher share of their income on food. Additionally, this issue is exacerbated by the local retailer market structure, as individuals in poorer areas have fewer alternatives and limited varieties of goods to substitute. Consequently, the combination of market segmentation and retailer market power creates a disproportionate burden on consumers in these regions. Therefore, policymakers should consider regional variations in inflation and market structure to alleviate the unequal impacts on economically disadvantaged areas.

Related Literature. This paper contributes to several major strands of the literature. First, our work relates to the literature on inflation heterogeneity across different groups. Hobijn and Lagakos (2005) and Hobijn et al. (2009) use Consumption Expenditure Survey (CEX) data to explore inflation differences between poor and nonpoor individuals. However, their analysis assumes that individuals have the same mix of goods within a category and face identical prices for a given good, varying only the expenditure shares of broad categories across people. Kaplan and Menzio (2015) and Kaplan and Schulhofer-Wohl (2017), using Nielsen Consumer Panel and Retail Scanner data, confirm inflation disparities across households with different income levels for the same bundle of goods and find that low-income and older households experience higher inflation on average. Jaravel (2018) finds similar results with the same data but emphasizes the role of product innovation and segmented consumption goods. He argues that increases in the relative demand for products consumed by high-income households led firms to introduce more new products and reduce the prices of continuing goods consumed by these households. Argente and Lee (2021) find that high-income households had lower inflation rates during the recession, as they were more able to substitute toward lower-quality goods. Handbury (2021) documents that welfare differences between rich and poor households may depend on the set of goods available in each region, with disparities growing in wealthy cities that offer the largest amenities. Outside of food inflation, Molloy (2024) documents heterogeneity in shelter inflation across income distributions. These studies primarily focus on inflation heterogeneity at the individual level or attribute it to consumer-related factors, such as differences in consumption baskets, price sensitivity, preferences, or search efforts. Our paper contributes to this body of work by offering new evidence of spatial heterogeneity in inflation across regions with different income level and identify retailer market structure as a novel source of this variation.

Another important strand of literature closely related to our paper is the growing body of work on retailer market concentration and market power. Pioneered by Autor et al. (2020) and De Loecker et al. (2020), numerous studies have documented the rising trends in market concentration and markups. Haltiwanger (2012) and Smith and Ocampo (2025) document that retailer market concentration has increased over time. In particular, Smith and Ocampo (2025) show that both national and local retail concentration levels have risen significantly, and the pattern is widespread. The expansion of multi-market retailers into new regions has been a key driver of this national concentration increase. Cao et al. (2024) find a similar trend of increasing concentration and highlight the rise of national chains, particularly dollar stores. Other studies, including Hottman (2017) and Stroebel and Vavra (2019), estimate retailer markups and explore their interaction with local characteristics.

Hottman (2017) finds that retailer markups are lower in large cities compared to smaller cities, while Stroebel and Vavra (2019) observe a positive correlation between retailer markups and local housing prices, suggesting that higher housing prices increase consumer wealth and lower the elasticity of consumption. On the other hand, Sangani (2022) documents that rich households pay significantly higher retail markups due to differences in search behavior. Our paper contributes to this literature by providing new evidence of retailer market concentration varying across regions with different income level. In particular, we document a novel finding that retailer market concentration is higher in lower-income areas and establish a causal relationship in which market concentration contributes to higher inflation rates in these areas.

Lastly, our study contributes to a broad set of studies examining the association between income inequality and price indexes. Contrary to our results, Moretti (2013) finds that real wage inequality is lower than nominal income inequality. This discrepancy may be due to differences across the studies in what goods are being measured and which areas are being considered.[6] Recent work (Martin, 2024) has also investigated the use of alternative price indexes that are not expenditure weighted. One concern with expenditure weighting is that the price indexes could be unrepresentative. Specifically, poor areas may contribute relatively less than rich areas to official price indexes given that poor areas consume less (even after adjusting for population). The poor areas further get down-weighted since we find that uniform pricing does not hold. The poorer areas are experiencing higher inflation, but the price of a given UPC is lower. This evidence runs contrary to some previous work by DellaVigna and Gentzkow (2019) that found uniform pricing within certain narrow categories within food.

The rest of the paper is organized as follows. Section 2 describes the data and key measures. Section 3 presents stylized facts on spatial heterogeneity in food inflation and retailer market structure. Section 4 outlines the empirical strategy to identify causality and presents the main findings. Section 5 discusses a potential mechanism through retailer market power and alternative hypotheses. Section 6 concludes the paper.

[6] We use a narrower set of goods but are broader in the areas considered, particularly in using more granular data.

2 Data and Measures

We use two main sources of data to analyze heterogeneous inflation rates across regions: the NielsenIQ Retail Scanner (RMS) dataset and Business Dynamic Statistics (BDS). The RMS dataset enables us to measure inflation rates and retailer market structure across regions by analyzing sales, price, and store distribution data from retailers for food products. The BDS dataset allows us to test the robustness of our findings by using alternative definitions of retailer size based on the number of employees.

2.1 NielsenIQ Retail Scanner

Our analysis uses the RMS dataset provided by the Kilts Center at Chicago Booth. This dataset includes weekly pricing, volume, and store merchandising data from over 100 retail chains across U.S. markets, covering more than 40,000 individual stores. Total sales in the NielsenIQ RMS sample exceed $200 billion annually, representing 50% of grocery store sales, 55% of drug store sales, 32% of mass merchandiser sales, and 2% of convenience store sales.

A key advantage of this dataset is that it contains detailed information at the finest product level, 12-digit universal product codes (UPCs) that uniquely identify specific goods. The dataset contains over 2.6 million UPCs. Furthermore, NielsenIQ classifies UPC-level goods by 10 departments, 110 product groups, and over 1,000 product modules. We further use a concordance provided by the U.S. Bureau of Labor Statistics (BLS) that maps NielsenIQ product modules to entry level items (ELIs).[7] These ELIs then map to Personal Consumption Expenditure (PCE) disaggregated categories.

Our analysis focuses on the food sector, which is identified as the aggregation of 21 PCE food categories, spanning from 2006Q1 to 2020Q4. Table 1 lists these 21 categories. The concordance between the PCE categories (based on the ELIs) and NielsenIQ product modules is derived from a mapping provided by the BLS, which links the ELIs to NielsenIQ product modules.

To construct our main dataset from NielsenIQ, we start with the weekly store-UPC-level raw data and link it to personal income data at the MSA level from the U.S. Bureau of Economic Analysis based on store location information in NielsenIQ.[8] We then define income deciles by the

[7] ELIs are the most granular complete mutually exclusive classification of CPI items produced by the BLS. We were provided this concordance as part of the Re-Engineering Statistics using Economic Transactions (RESET) project.

[8] Note that our baseline analysis relies on the MSA location of retailer stores in NielsenIQ. Potential concerns

Table 1: 21 PCE Food Categories

1	Bakery	12	Milk
2	Beef and Veal	13	Other Foods
3	Beer	14	Other Meats
4	Cereal	15	Pork
5	Coffee	16	Poultry
6	Dairy	17	Processed Fruits and Vegetables
7	Eggs	18	Soda
8	Fats and Oil	19	Spirits
9	Fish	20	Sugar and Sweets
10	Fruits	21	Vegetables
11	Wine		

Notes: The table represents the 21 PCE disaggregated Food categories. These disaggregated categories are mutually exclusive. The PCE category Food and Beverages is composed of these 21 categories.

Table 2: Examples of MSA Deciles

Decile 1	El Paso (TX), Albany (GA), Yuma (AZ), Terre Haute (IN), etc.
Decile 5	Knoxville (TN), Panama City (FL), Binghamton (NY), Wilmington (NC), etc.
Decile 10	New York (NY), Washington (DC), Boston (MA), San Francisco (CA), etc.

Note: The table provides some examples of MSAs located in the decile 1, 5, and 10. These deciles are time invariant in our setting and are based on income per capita data from the BEA.

cross-time average of MSA-level income per capita. See examples of the income deciles in Table 2. The data is aggregated to monthly frequency using the National Retail Federation (NRF) calendar and then up to the quarterly level.[9] Using the concordance between product modules and PCE food categories, we identify the food sector in NielsenIQ. Finally, to measure retailer market structure and the degree of competition, we link store identifiers to retail chain identifiers using the crosswalk provided by Nielsen.

Our main analysis is at the MSA (or MSA income decile), food category, and quarter level. We generate price indexes, the Herfindahl-Hirschman index, and other statistics associated with market power and structure for each pairing of MSA (or MSA income decile) and food category-quarter.

about this measure arise if an MSA is broad enough to encompass consumers who move across MSAs, potentially creating a gap between consumer income and that of residents. To address this, we leverage the Nielsen Consumer Panel data to examine the fraction of households shopping outside their residential MSAs and explore their characteristics. Additionally, we compare two definitions of income deciles, one based on consumer MSAs and the other based on household MSAs. More details are provided in Appendix A, which help address potential concerns.

[9]The NRF calendar typically starts in early February and ends around the end of January in the following year.

Table 3: Summary Statistics of MSA-quarter level Sample

	Mean (SD)
Income per capita ($ thousands)	42.45
	(9.30)
Sales ($ millions)	206.67
	(365.60)
Number of chains	9.75
	(3.72)
Number of stores	193.20
	(250.78)
Number of UPCs	49177.23
	(18953.68)
Share of large chains	0.08
	(0.12)
Share of small chains	0.10
	(0.11)
Market Concentration	0.31
	(0.15)
Observations	11,168
Number of MSAs	188
Number of quarters	60

Note: The table provides the summary statistics of the main MSA-level sample for the aggregate food and beverages. Large (small) chains are defined by those in the top (bottom) decile based on the count of stores of that chain at the national level within an MSA-quarter. Market concentration is measured by the HHI (Herfindahl-Hirschman Index) of chain-level sales.

We use HHI as our main measure of market concentration by relying on data from the NielsenIQ Retailer Scanner dataset. Table 3 provides summary statistics for the main sample.

2.2 Business Dynamic Statistics

The Business Dynamic Statistics (BDS, henceforth) is a public version of administrative Census firm-level data, the Longitudinal Business Dynamics. The data provide annual measures of business dynamics in the U.S., such as job creation and destruction, establishment births and deaths, and firm entry and exit. These data are provided for the economy overall as well as aggregated by establishment or firm characteristics such as firm size and age. Furthermore, the data provide sectoral- and geographic-level information, which allows us to track the business dynamics at the sector, state, county, and MSA levels.[10] In the BDS, we use retailers' information (based on NAICS code 44-45) and construct alternative measures for retailer size and market structure at the MSA level.

2.3 Main Measures

2.3.1 Price Indexes

To measure and compare the cost of living across income deciles, we construct price indexes from the UPC-level data in NielsenIQ. As a starting point, we use traditional price indexes, focusing on the log geometric Laspeyres price index, which is calculated as follows:

$$\ln \Psi_{mt}^{G} = \sum_{k \in \mathbb{C}_{mt-1,mt}} \omega_{mkt} \ln \frac{p_{mkt}}{p_{mkt-1}}, \tag{1}$$

where ω_{mkt} represents the weight assigned to product k in quarter t for MSA m, and we use lagged expenditure shares as weights ($\omega_{mkt} = s_{mkt-1}$) for the Laspeyres index. The set $\mathbb{C}_{mt-1,mt}$ consists of all "continuing" goods that are sold in both periods t and $t-1$ in MSA m.

Although our default measure is the geometric Laspeyres index, we also use the geometric Paasche index, which replaces the weights with current expenditure share ($\omega_{mkt} = s_{mkt}$). Additionally, we conduct a robustness test using alternative demand-based indexes based on the constant elasticity of substitution (CES) preference assumption, to account for potential substitution bias inherent in the traditional indexes.[11] One is the Sato-Vartia index, where we replace the above weight with

$$\omega_{kt} = \frac{\frac{(s_{kt}-s_{kt-1})}{(\ln s_{kt}-\ln s_{kt-1})}}{\sum_{k\in\mathbb{C}_{t-1,t}} \frac{(s_{kt}-s_{kt-1})}{(\ln s_{kt}-\ln s_{kt-1})}},$$

which accounts for product entry and exit, in addition to the demand effects for common goods appearing between periods $(t-1)$ and t. Another index we use is the Feenstra-adjusted Sato-Vartia

[10] See more details in https://www.census.gov/programs-surveys/bds.html.

[11] The traditional indexes do not account for demand effects that may arise from consumers substituting between differentiated goods.

index, which incorporates the effects of product entry and exit. It is constructed using the following formula:

$$\ln \Psi_{mt}^{Feenstra-SV} = \ln \Psi_{mt}^{SV} + \frac{1}{\sigma - 1} \ln \frac{\lambda_{mt,t-1}}{\lambda_{mt-1,t}},$$

where $\lambda_{mt,t-1} = \frac{\sum_{k \in \mathbb{C}_{mt-1,t}} p_{mkt} q_{mkt}}{\sum_{k \in \Omega_{mt}} p_{mkt} q_{mkt}}$, $\lambda_{mt-1,t} = \frac{\sum_{k \in \mathbb{C}_{mt-1,t}} p_{mkt-1} q_{mkt-1}}{\sum_{k \in \Omega_{mt-1}} p_{mkt-1} q_{mkt-1}}$.

Lastly, we also construct the price indexes by restricting our sample to UPCs sold in all ten income deciles in a given quarter. Consumption baskets vary across different income groups, as indicated in Jaravel (2018), and potentially across regions with different income levels. Therefore, we use this price index, based only on the set of common goods, to assess whether regional inflation disparities stem from the differences in consumption baskets. Our findings show that applying the common goods restriction reduces the inflation gap between regions, but it does not fully explain the difference in inflation between the top and bottom income deciles.

2.3.2 Retailer Market Structure

In the Nielsen IQ data, we define large and small chains based on the distribution of store counts at the national level. Using store and retailer codes along with geographic information for each store, we identify stores, retailers, and their ownership structures across regions and time. We define the size of retailers based on the number of stores they own at the national level. We classify large chains as those in the top decile and small chains as those in the bottom decile of the size distribution. We then calculate the number and share of large and small chains in each MSA.

Alternatively, using the BDS, we define large and small retailers based on their number of employees. Large retailers are those with 500 or more employees, while small retailers have 19 or fewer employees. We then calculate the share and employment share of large and small firms within each MSA and compare these across different regions.

Finally, we use the sales share of retailers to construct the Herfindahl-Hirschman index (HHI) to measure the degree of market concentration among retailers in each region.

3 Spatial Heterogeneity in Inflation and Retailer Market Structure

3.1 Price and Inflation Patterns

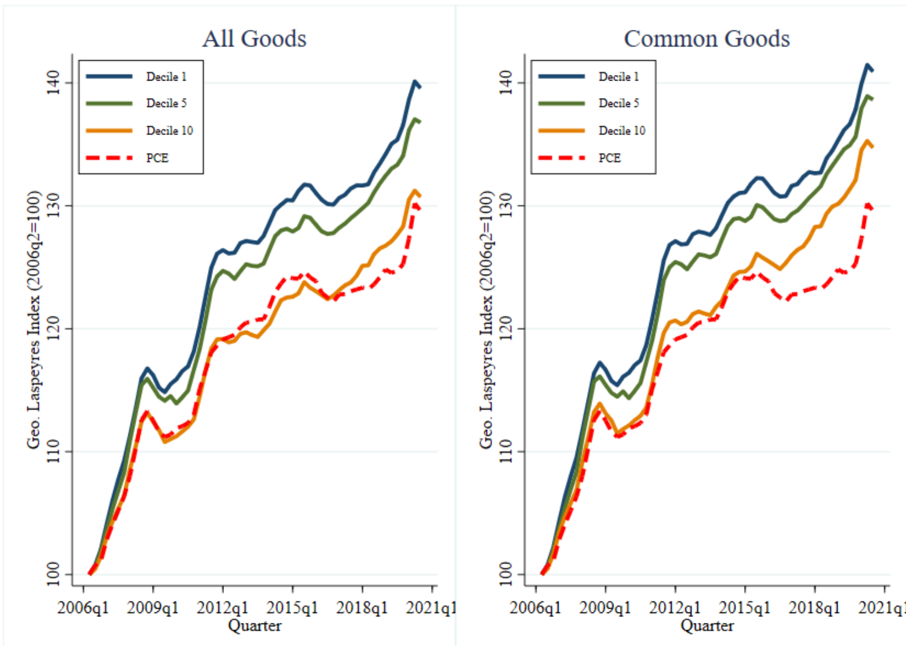


Figure 1: Price Index for Aggregated Food

Notes: This figure represents relative prices for the aggregated food market with four series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2020Q4. The data for the three solid lines come from the NielsenIQ Retail Scanner dataset, represented by geometric Laspeyres price indexes, while the dashed line is the official measure of personal consumption expenditures (PCEs) from the U.S. Bureau of Economic Analysis (official measure). Each solid line corresponds to a decile of the income per capita ranking of MSAs, with decile 1 containing the MSAs with the lowest income per capita and decile 10 containing the MSAs with the highest income per capita. The left panel shows results for the set of goods sold by retailers in quarters t and t-1. The right panel corresponds to the set of goods present across all 10 deciles in quarters t and t-1. We map the NielsenIQ UPCs to the PCE definition of food purchased for off-premises consumption by using a product module concordance provided by the U.S. Bureau of Labor Statistics.

Figure 1 presents the geometric Laspeyres index constructed from the NielsenIQ Scanner data, alongside the official PCE price index for the first (poorest), fifth, and tenth (richest) income deciles. The analysis focuses on aggregated food. The left panel shows the price index including all UPCs, while the right panel includes only common goods—those UPCs present across all deciles. The base quarter is set to 2006Q2.[12]

The general trend captured by Figure 1 indicates that the poorest decile ("Decile 1") exhibits higher price growth compared to the richer deciles ("Decile 5" and "Decile 10"). This pattern holds even when restricting the sample to the set of common goods for the price index construction. These results suggest that the variation in price growth across deciles is not primarily driven by different consumption baskets or preferences among consumers in different regions. This trend is generally consistent across the 21 PCE food categories as well as for other aggregated food series. Furthermore, these patterns remain robust even after using demand-based price indexes. See Appendix B for further details.

Lastly, note that the official PCE series is closer to the series for the highest income decile than it is to any other decile. This demonstrates that the official PCE price index series understates inflation to the largest extent for individuals living in the poorest areas. This discrepancy in inflation has significant macroeconomic implications. For example, if we assumed uniform nominal wage growth across the United States, the official real wage growth would appear higher than the actual real wage growth experienced in the poorest areas.

3.2 Retailer market structure

To examine retailer market structure across different regions, we compute summary statistics for our main sample from NielsenIQ by income-per-capita decile. See Table 4. The table shows that richer income areas have a greater number of retailers and stores as well as higher sales. In addition, the share of large chains is higher, while the share of small chains is lower in poorer income regions. Also, poorer areas have fewer UPCs and have higher quantity and expenditure shares of total consumption allocated to the set of common goods. Finally, poorer income areas experience a higher degree of sales market concentration.

Moreover, Figure 2 displays the distribution of retailers of different size (the national store count) across income deciles. In line with Table 4, this shows that the poorer deciles have a smaller mass of small retailers (with fewer stores at the national level) relative to the richer deciles.

We next run the following regressions to examine the cross-sectional variation in retailer market

[12] Note that price indices are constructed using information from both periods t and t-1. Thus, 2006Q2 is the first quarter in which we are able to estimate a price index.

Table 4: Summary Statistics of MSA-quarter level Sample by Income Deciles

	Decile 1	Decile 3	Decile 5	Decile 7	Decile 10
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Income per capita ($ thousands)	29.660	34.438	36.843	39.760	52.424
	(2.661)	(0.442)	(0.384)	(0.505)	(6.675)
Sales ($ millions)	27.483	68.282	111.430	136.036	852.749
	(23.213)	(86.639)	(240.192)	(139.703)	(770.987)
Population Share	0.024	0.038	0.057	0.060	0.378
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
Number of chains	7.334	8.494	8.904	9.607	13.779
	(2.473)	(2.698)	(3.548)	(3.069)	(4.971)
Number of Stores	59.296	97.954	125.597	140.596	571.255
	(55.378)	(124.141)	(169.693)	(104.836)	(502.549)
Share of Large Chains	0.14	0.12	0.11	0.11	0.07
	(0.07)	(0.06)	(0.06)	(0.06)	(0.04)
Share of Small Chains	0.05	0.07	0.06	0.06	0.16
	(0.07)	(0.09)	(0.09)	(0.07)	(0.13)
Number of UPCs (thousands)	34.772	41.745	44.011	49.385	72.259
	(11.308)	(14.096)	(17.196)	(16.650)	(22.878)
Market Concentration	0.397	0.359	0.346	0.319	0.176
	(0.164)	(0.144)	(0.140)	(0.128)	(0.085)

Note: The table provides the summary statistics of the main MSA-level sample for the aggregate food and beverages for the five income-per-capita deciles 1, 3, 5, 7, 10. Large (small) chains are defined by those in the top (bottom) decile based on the count of stores of that chain at the national level within an MSA-quarter. Market concentration is measublack by the HHI (Herfindahl-Hirschman Index) of chain-level sales.

structure across MSAs with different income levels:

$$Y_{mt} = \beta_0 + \beta_1 Income_{mt} + \delta_m + \delta_t + \varepsilon_{mt},$$

where Y_{mt} is either the sales, total count of chains or stores, or the share of large retailers in MSA m in quarter t. $Income_{mt}$ is income per capita in MSA m, and δ_m and δ_t are MSA and quarter fixed effects, respectively. The results, presented in Table 6, confirm the cross-sectional patterns that richer areas have higher sales, more retailers and stores, and a lower fraction of large retailers.

We observe consistent patterns in the BDS data as well. See Appendix C for further details.[13]

[13]Figure C.1 shows that more retail chains are located in richer areas, and Figure C.2 shows that these retailers create more jobs in those areas. Furthermore, we find a clear pattern between firm size and income decile. Figures C.3 and C.4 present the share of large and small retailers, respectively, within each income decile.

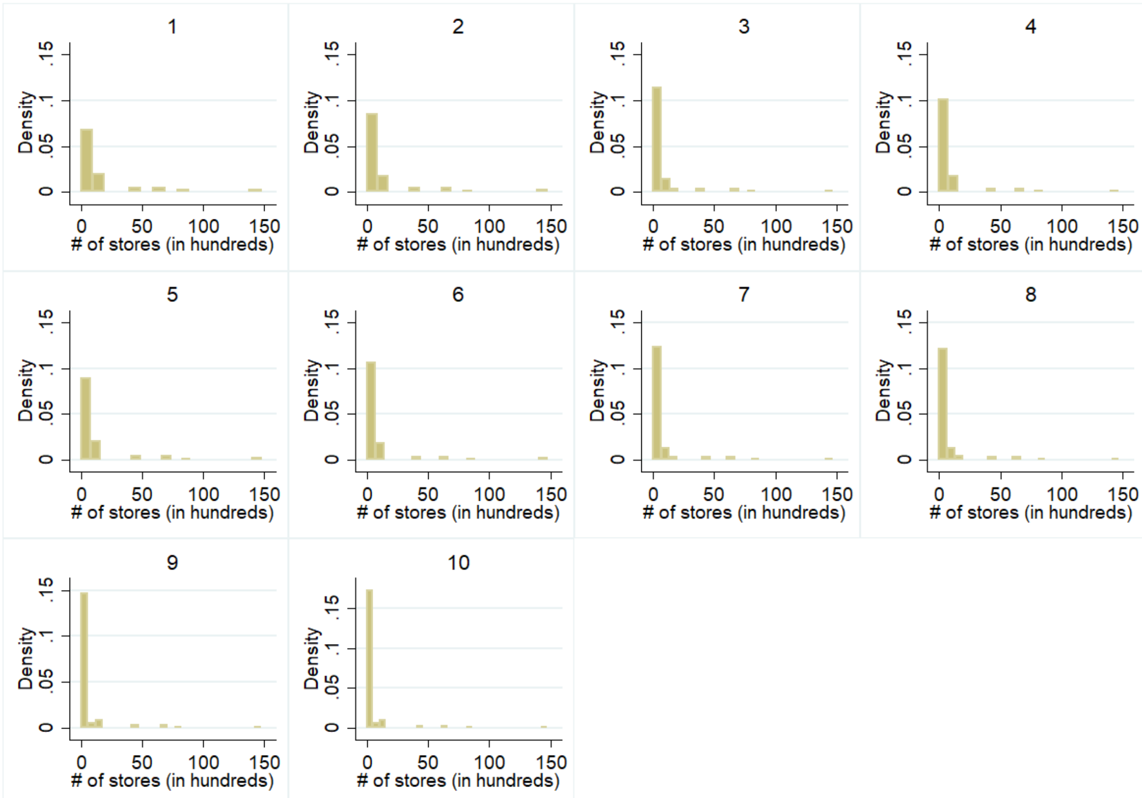


Figure 2: Distribution of Retailer Size by Income Decile
Notes: The figure shows the distribution of the number of stores (at the national level) of chains located in each income decile. The data come from NielsenIQ for the aggregate food and beverages.

These figures indicate that poorer income areas have a larger share of large retailers, while richer areas tend to have a smaller share of large retailers. Conversely, the fraction of small retailers is higher in wealthier areas.[14]

We next explore retailer market concentration at the decile level using the following regression:

$$HHI_{idt} = \beta_0 + \beta_1 Decile_{dt} + \delta_i + \delta_t + \varepsilon_{idt},$$

where HHI_{idt} is the Herfindahl–Hirschman index of retailer sales for the PCE food category i in MSAs in income decile d in quarter t. $Decile_{dt}$ is an indicator for income decile, and δ_i and δ_t

[14]Note that the size of retailers is measured by firm-level employment, and the share is calculated based on the number of firms operating retail stores in each MSA. This analysis is robust to using the number of establishments instead. Additionally, these patterns hold consistently across the entire sample period.

Table 5: Retailer Dynamics in NielsenIQ

	Sales (in \$1mil.)	Chain counts	Store counts	Large firm share
Income	14.01***	0.108***	2.770***	-0.049**
	(0.422)	(0.007)	(0.192)	(0.021)
Quarter FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
Observations	11,168	11,168	11,168	11,168

Note: The table represents regression results from our two-way fixed effects estimator. The coefficient of interest is the coefficient on income per capita (in \$1000) in an MSA for a given quarter. The dependent variable is total sales in in Column 1, total counts of chains and stores in in Column 2, 3, and an unweighted share (%) of large firms in Column 4, where large retailers are defined by the top decile of the number of store counted at the national level in NielsenIQ. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Retailer Dynamics in NielsenIQ

	Sales (in \$1mil.)	Chain counts	Store counts	Large firm share
Income	14.01***	0.108***	2.770***	-0.049**
	(0.422)	(0.007)	(0.192)	(0.021)
Quarter FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
Observations	11,168	11,168	11,168	11,168

Note: The table represents regression results from our two-way fixed effects estimator. The coefficient of interest is the coefficient on income per capita (in \$1000) in an MSA for a given quarter. The dependent variable is total sales in in Column 1, total counts of chains and stores in in Column 2, 3, and an unweighted share (%) of large firms in Column 4, where large retailers are defined by the top decile of the number of store counted at the national level in NielsenIQ. *** p<0.01, ** p<0.05, * p<0.1

represent fixed effects for the PCE food category and time, respectively. The results in Table ?? show that the HHI is higher in lower income deciles, indicating that poorer deciles experience a higher degree of market concentration.

We also estimate consumers' elasticity of substitution between products, following Feenstra (1994), to understand how consumption behavior is correlated with retailer market dynamics and the pricing dispersion observed across different regions. Table 8 presents the cross-time average of the HHI and the elasticity of substitution for a subset of 21 food items and aggregated food.

These results suggest that retailer market structure varies across regions with different income levels. In particular, retailer market concentration is higher in poorer areas, where a larger share of sales is dominated by larger firms. Additionally, poor income areas tend to exhibit a lower elasticity of substitution with fewer varieties, stores and chains. This indicates that consumers in these regions are less likely to substitute different goods into their baskets.

Table 7: HHI across Different Income Deciles

	HHI
Decile	-0.004***
	(0.000)
Year FE	Yes
Item FE	Yes
Observations	10,920

Note: The table represents regression results from our two-way fixed effects estimator. The coefficient of interest is the coefficient on income per capita decile. This independent variable is a discrete categorical variable that takes the value 1 (poorest) to 10 (richest). The outcome variable is HHI. HHI is the Herfindahl-Hirschman index (HHI) of retail chain's sales of a given PCE disaggregated category within an MSA. HHI is a continuous variable than can range from 0 to 1. Column 1 is an unweighted share of large firms. Column 2 is an employment weighted large share of retailers. Data is collected from the Business Dynamics Statistics and retailers are gathered from retail trade sector (NAICS 44-45). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

4 Retailer Market Concentration and Inflation Disparities

To explore the potential link between higher inflation and retailer market concentration in poorer MSAs compared to richer ones, we conduct further analyses using the Herfindahl-Hirschman Index (HHI) as our measure of market concentration. First, we examine the relationship between inflation rates and HHI. To assess a causal relationship, we exploit a quasi-experiment based on the 2015 bird flu outbreak and apply a triple-difference estimator. Consequently, this section focuses on egg price inflation.

Table 8: HHI and the Elasticity of Substitution

Item	Decile	HHI	Elasticity of substitution
Cereal	1-3	0.1337	4.3106
Cereal	4-7	0.1343	4.3289
Cereal	8-10	0.1361	5.4731
Eggs	1-3	0.3324	3.5803
Eggs	4-7	0.3104	7.6531
Eggs	8-10	0.2892	8.1946
Fats and Oil	1-3	0.0639	4.0299
Fats and Oil	4-7	0.0610	4.1235
Fats and Oil	8-10	0.0580	4.6153
(Alcoholic Beverages)			
Beer	1-3	0.2798	6.2084
Beer	4-7	0.2311	6.5024
Beer	8-10	0.1740	8.1054
Spirits	1-3	0.0515	5.3730
Spirits	4-7	0.0492	6.2725
Spirits	8-10	0.0474	7.0815

Note: Each subpanel represents one of the 21 PCE-items with statistics on HHI when calculated using all goods. We show the average of three subgroups based on deciles of the income per capital ranking of MSAs: the average of deciles 1-3 (three lowest income per capita deciles), the average of deciles 4-7 (median income per capita deciles), and the average of deciles 8-10 (three richest income per capita deciles). The HHI measures levels of market concentration with a range of 0 to 1 where values closer to 1 represent higher levels of market concentration. All of the statistics are produced using the NielseIQ Retail Scanner dataset, averaged over 2006Q1-2016Q4. The elasticity of substitution is constructed following the method in Feenstra (1994). The elasticity of substitution measures how easy it is for individuals in those deciles to substitute across goods in the corresponding local market where higher values correspond to higher ease of substitution. Note that the last two items are alcoholic beverages, which belong to the broadest aggregate foods category named “Food and Beverages”.

4.1 Standard OLS Estimator

First, we test how the inflation rate at the MSA level is associated with the degree of market concentration using the following simple OLS regression:

$$P_{mt} = \beta_0 + \beta_1 HHI_{mt} + \delta_m + \delta_t + \varepsilon_{mt}, \tag{2}$$

where P_{mt} is the (geometric) Laspeyres inflation rate of eggs in MSA m in quarter t. HHI_{mt} is the HHI of retailer sales in MSA m in quarter t. δ_m and δ_t are the MSA and quarter fixed effects,

Table 9: Market Concentration in Eggs Market

	Inflation
HHI	0.022***
	(0.005)
Constant	-0.003
	(0.003)
Observations	9,484

Note: The table represents regression results from our two-way fixed effects estimator. The coefficient of interest is the coefficient on our measure of market concentration: HHI. HHI is the Herfindahl-Hirschman index (HHI) of retail chain's sales of eggs within an MSA. HHI is a continuous variable than can range from 0 to 1. The dependent variable is inflation at the MSA-quarter level. Inflation is measured using the geometric Laspeyres price index. HHI and inflation measures are based on NielsenIQ Retail Scanner data. *** p<0.01, ** p<0.05, * p<0.1

respectively.

The results are presented in Table 9, where we find a positive and statistically significant relationship between HHI and inflation.[15] However, it is important to note that we cannot establish causal relationships here, as this analysis may be subject to endogeneity bias. For instance, the observed relationship could be demand-driven, where consumers in MSAs with higher HHI are more likely to purchase goods that are experiencing relatively higher inflation. Alternatively, consumers in richer MSAs may differ significantly from those in poorer MSAs, with greater sensitivity to price changes. Such heterogeneity in consumer behavior may have led retailers in wealthier areas to increase prices at slower rates. Another potential explanation is a supply-side story, where poorer MSAs have fewer stores, which weakens competition and allows retailers to raise prices.

To isolate whether the effect we observe is driven by the supply side or the demand side, we use the 2014–2015 bird flu outbreak as a quasi-experiment. In the following section, we apply a triple-difference estimator to investigate this relationship in greater detail.

[15] In this specification, and all other specifications with MSA-level data, we cluster standard errors at the MSA level.

4.2 Triple-Difference Estimator

We use the 2014-2015 highly pathogenic avian influenza outbreak as an exogenous supply shock to the egg market. The 2015 bird flu episode affected the price and quantities of eggs sold, as evidenced in Figure B.1 around 2014Q4–2015Q1. Based on U.S. Department of Agriculture (USDA) reports, 36 million layers (birds that lay eggs) were lost due to the bird flu by June 2015.[16]

Importantly, reports from the USDA and the Government Accountability Office (GAO) indicate that the impact of the bird flu shock exhibited geospatial variations, primarily affecting the central and western parts of the U.S. We have access to official data from the USDA on the timing, locations, and number of layers that were culled.[17] By identifying the MSAs where these layers were culled, we can pinpoint areas disproportionately affected by the bird flu, which may have experienced higher inflation in egg prices early in the outbreak during the inflationary period.

Leveraging this information, we can construct a difference-in-differences identification strategy by grouping treated and control MSAs and comparing the effect of the bird flu outbreak on them. Additionally, we extend this approach by using a triple difference-in-differences estimator, which interacts MSA-level market concentration with the standard diff-in-diff term. This allows us to examine how the effect of the outbreak varies based on the degree of retailer market concentration.

First, to measure whether the MSAs where farmers culled their layers were disproportionately affected by the bird flu, we use a two-year window around the treatment in 2014Q4 and run the following traditional two-way fixed effects regression over the sample from 2012Q4 to 2016Q4:

$$P_{mt} = \beta_0 + \beta_1(Treated_m \times Post_t) + \delta_m + \delta_t + \varepsilon_{st}, \tag{3}$$

where P_{mt} is the (geometric) Laspeyres inflation rate for eggs in MSA m in quarter t. $Treated_m$ is an indicator variable that takes the value of one if farmers in MSA m had to cull their layers during the 2014-2015 bird flu outbreak, according to the USDA. $Post_t$ is a binary variable that takes the value of one after 2014Q4, and zero otherwise. As before, δ_m and δ_t are the MSA fixed effects and quarter fixed effects, respectively. The coefficient on β_1 should be positive, at least during the inflationary period of the bird flu, given that these MSAs experienced a relatively larger cost shock.

[16]The USDA also compensated producers that had to cull their layers. Payment was based on ”fair market” values as determined by USDA appraisers.

[17]https://crsreports.congress.gov/product/pdf/R/R44114

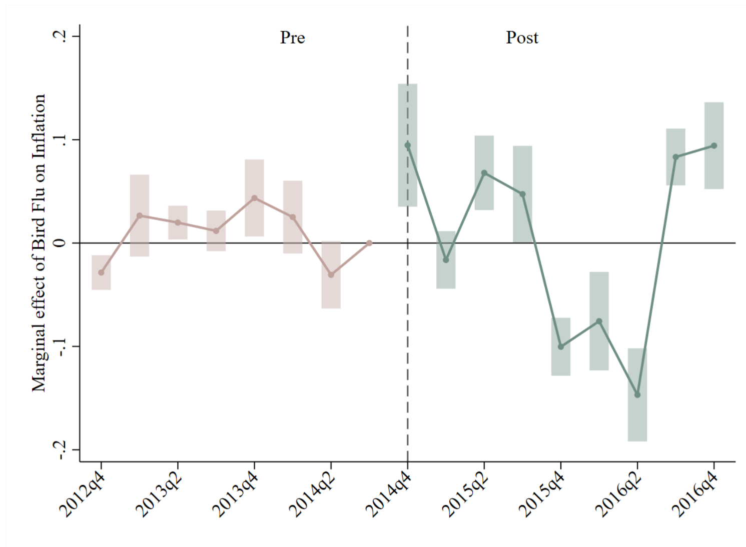
Table 10: TWFE Estimator (Bird Flu Episode)

	Inflation	Inflation	Inflation	Abs. Inflation
Bird Flu × Post	-0.003	0.039***	-0.035***	0.053***
	(0.004)	(0.008)	(0.007)	(0.006)
Sample	All	Inflation	Deflation	All
Quarter FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
Quarters	17	10	7	17
MSAs	187	187	187	187
Observations	3,160	1,859	1,301	3,160

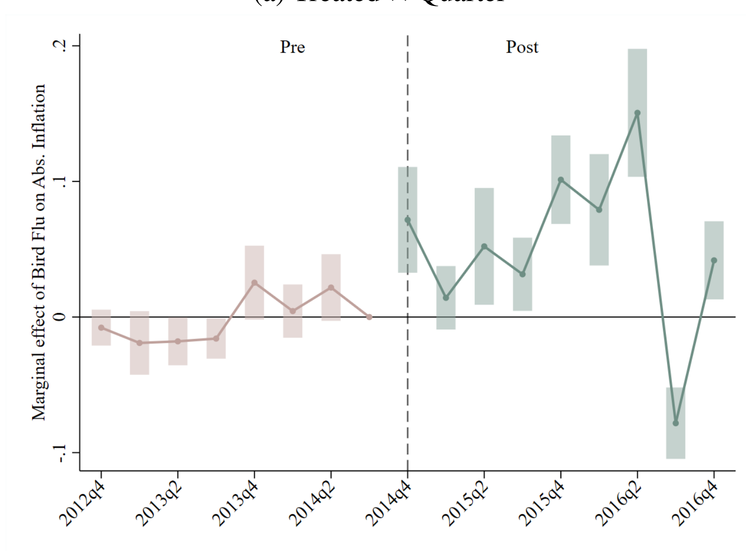
Note: The table represents regression results from our two-way fixed effects estimator. The coefficient of interest is the interaction of Post and Bird Flu. Post us a binary variable that takes the value 1 after 2014q4. Bird Flu is a binary variable that takes the value 1 if an MSA culled its layers during the 2014-2105 bird flu episode. Column 1 pools all periods together and has the inflation rate as the outcome variable. Column 2 only looks at the inflationary period and has the inflation rate as the outcome variable. Column 3 only looks at the deflationary period and has the inflation rate as the outcome variable. Column 4 pools all periods together and has the absolute value of the inflation rate as the outcome variable. Inflationary and deflationary periods are determined by the national price index of eggs. The sample period ranges from 2012q4 to 20164. MSA and quarter fixed effects are included. Standard errors are clustered at the state-level. *** p<0.01, ** p<0.05, * p<0.1

The results are shown in Table 10. In column 1, we estimate an effect of zero, which may suggest that these MSAs were not disproportionately affected by the 2014-2015 bird flu. However, this null effect masks heterogeneity in the effects during this period. When we separate the sample into inflationary and deflationary periods, we observe opposing effects in the MSAs where layers were culled. In column 2, we restrict the sample to the inflationary period when the national egg inflation rate was above zero. Here, we estimate a 0.04 coefficient on the interaction of Bird Flu and Post, which corresponds to a 4 percentage point higher inflation rate in MSAs affected by the bird flu after 2014Q4. This point estimate is significant at the 1% level. In column 3, we restrict the sample to the deflationary period and find that MSAs that culled their layers experienced a 4 percentage point lower inflation rate after 2014Q4. This point estimate is also significant at the 1% level. In column 4, we pool all quarters and take the absolute value of the dependent variable (the inflation rate). We find that MSAs that culled their layers experienced larger changes in the inflation rate after 2014Q4.

These heterogeneous inflation effects during the bird flu outbreak are reflected in Figure 3. The left panel plots the standard event study difference-in-differences coefficients. The dashed



(a) Treated × Quarter



(b) Treated × Quarter (Absolute Value)

Figure 3: Event Study Difference-in-Differences (Bird Flu)

Notes: The figure represents the event study difference-in-differences analysis examining the dynamic effect of MSAs disproportionately affected by the 2014-2015 bird flu episode on inflation. The outcome variables in the left and right panels are, respectively, the inflation rate and the absolute value of the inflation rate. MSAs are assigned to the treatment group based on a USDA report detailing which farms culled their layers. The post-period starts in 2012q4, and 2012q3 is the reference quarter. Effects are measured from 2012q4 to 2016q4. Standard errors are clustered at the MSA level.

vertical line corresponds to 2014Q4, marking the start of the post-period. We observe no systematic difference in inflation rates between MSAs that culled their layers (the treated MSAs) and MSAs that did not cull their layers (the control MSAs) prior to 2014Q4. This satisfies the parallel trends assumption. However, after 2014Q4, the treated MSAs experienced relatively higher inflation in some quarters and relatively lower inflation in others, suggesting a mixed but significant effect of the bird flu shock.

These opposing inflation effects can be explained by heterogeneous impacts depending on whether the egg market is in an inflationary or deflationary period. This dependency on the inflationary or deflationary phase is reflected in the right panel of the figure, where we replace the dependent variable with the absolute value of the inflation rate. We find that the impacted MSAs were consistently more affected after 2014Q4. Additionally, we continue to observe no significant difference between these two groups of MSAs prior to 2014Q4, further supporting the notion that the bird flu shock had a heterogeneous effect depending on market conditions.

Next, we use a triple-difference estimator to measure how the impact varies across the treated MSAs (where farmers culled their layers) with different degrees of market concentration (HHI). The following regression outlines the identification strategy:

$$\begin{aligned} P_{mt} = \beta_0 &+ \beta_1 HHI_{mt} + \beta_2(\text{Treated}_m \times \text{Post}_t) \\ &+ \beta_3(\text{Treated}_m \times HHI_{mt}) + \beta_4(\text{Post}_t \times HHI_{mt}) \\ &+ \beta_5(\text{Treated}_m \times \text{Post}_t \times HHI_{mt}) + \delta_m + \delta_t + \varepsilon_{mt}, \end{aligned} \tag{4}$$

where the subscript m corresponds to MSA m and t is quarter t. $Treated_s$ is a binary variable indicating whether MSA s is near to where layers were culled during the 2014-2015 bird flu episode according to the USDA report. $Post_t$ is a binary variable that takes the value 1 if quarter t is after 2014Q4, and zero otherwise. HHI_{mt} is the HHI of retailer concentration of sales in MSA m for quarter t. P_{mt} is the geometric Laspeyres inflation rate in MSA m in quarter t. The fixed effect terms, δ_m and δ_t, are the same as before, and ε_{st} is the error term.

The results are presented in Table 10. In column 1, we restrict our sample to the inflationary period and find that MSAs with higher market concentration experienced faster price increases in the egg market after the bird flu episode. This point estimate is significant at the 1% level. One concern is that these MSAs with higher market concentration might potentially lower prices by larger amounts during the deflationary period. We find no support for this hypothesis when we restrict our sample to the deflationary period in column 2. In fact, we find some suggestive evidence that these MSAs are slower to decrease prices in the deflationary period, as indicated by the positive

Table 11: Triple Difference Estimator (Market Concentration)

	Inflation	Inflation	Inflation
Bird Flu × Post × HHI	0.078***	0.065*	0.052***
	(0.021)	(0.038)	(0.018)
Bird Flu × Post	-0.002	-0.078***	-0.034**
	(0.012)	(0.026)	(0.014)
HHI × Post	-0.003	-0.006	-0.006
	(0.008)	(0.012)	(0.005)
Bird Flu × HHI	-0.180	1.246*	0.389
	(0.137)	(0.656)	(0.297)
HHI	0.073***	0.022	0.050***
	(0.019)	(0.037)	(0.018)
Sample	Inflation	Deflation	All
Quarter FE	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes
Quarters	10	7	17
MSAs	187	187	187
Observations	1,859	1,301	3,160

Note: The table represents regression results from our triple difference-in-differences. The coefficient of interest is the interaction of Post, HHI, and Bird Flu. Post us a binary variable that takes the value 1 after 2014q4. HHI is the Herfindahl-Hirschman index (HHI) of retail chain's sales of eggs within an MSA. HHI is a continuous variable than can range from 0 to 1. Bird Flu is a binary variable that takes the value 1 if an MSA culled its layers during the 2014-2105 bird flu episode. Column 1 only looks at the inflationary period. Column 2 only looks at the deflationary period. Columns 3 pools all periods together. Inflationary and deflationary periods are determined by the national price index of eggs. The sample period ranges from 2012q4 to 20164. MSA and quarter fixed effects are included. Standard errors are clustered at the state-level. *** p<0.01, ** p<0.05, * p<0.1

coefficient on the triple interaction term in column 2. This coefficient is significant at the 10% level. In column 3, we pool all quarters from the two-year window together and continue to find that MSAs with higher market concentration exhibit higher inflation than those with lower market concentration.

Our results indicate that retailer market concentration contributes to the heterogeneous inflation rates between poor and rich MSAs in the egg market. A potential mechanism behind the role of retailer market concentration is their market power. In particular, the triple difference-in-differences results suggest that these differences may stem from variations in markups rather than marginal costs with high cost pass-through. If differences in marginal costs were the primary driver behind the heterogeneous inflation rates, we would expect to observe greater deflation in the deflationary period in MSAs with higher market concentration. To explore this connection, we develop a simple model in the next section, where retailer markups are determined by their market share. Furthermore, we plan to expand on these analyses will estimate markups and test this hypothesis directly to disentangle the sources of the observed effect.

5 Potential Mechanism through Market Power

A potential mechanism behind the role of retailer market concentration is the market power held by retailers. To explore this connection, we develop a model based on the nested CES structure and Bertrand competition among retailers. In this framework, retailer markups are determined by their market share, which is influenced by the degree of concentration in the market. Specifically, higher market concentration can lead to reduced competitive pressure, allowing retailers to raise their markups. This model helps to explain why MSAs with higher market concentration may experience more pronounced inflationary effects during periods of shock.

5.1 Theoretical Framework

5.1.1 Consumer Preferences

Suppose that consumers consume a variety of goods from multiple stores. In the first stage, they choose which retail store to shop from based on the retailer price indices. In the second stage, once a store is selected, consumers decide which food categories (e.g., eggs, milk, etc.) to purchase, guided by the price indices of these food items. In the final stage, within a chosen store and food category, consumers select a specific UPC (barcode item, e.g., 8 oz. Almond Milk) based on its price. The demand of the representative consumer follows a standard nested CES demand structure.

Given this, the utility of the presentative consumer in MSA m at time t is assumed to be:

$$U_{mt} = \left[\sum_{s \in S_{mt}} (\varphi_{smt} C_{smt})^{\frac{\sigma_S - 1}{\sigma_S}} \right]^{\frac{\sigma_S}{\sigma_S - 1}}, \tag{5}$$

where C_{smt} is the consumption index of store s in MSA m at time t; φ_{smt} is the quality of store s at time t; S_{mt} is the set of stores in MSA m at time t; and σ_S is the constant elasticity of substitution across stores within the MSA.

The consumption index C_{smt} is itself a CES aggregator of the consumption indices for food item i (among the 21 PCE food items) from store s at time t, as follows:

$$C_{smt} = \left[\sum_{i \in I_{smt}} \left(\varphi_{ismt} C_{ismt} \right)^{\frac{\sigma_I - 1}{\sigma_I}} \right]^{\frac{\sigma_I}{\sigma_I - 1}}, \tag{6}$$

where φ_{ismt} is the quality of food item i at store s at time t; I_{smt} is the set of food items sold by store s at time t; and σ_I is the constant elasticity of substitution across food items within the store.

The consumption index for each food item, C_{ismt}, is also a CES aggregator and is given by:

$$C_{ismt} = \left[\sum_{u \in U_{ismt}} \left(\varphi_{usmt} C_{usmt} \right)^{\frac{\sigma_U - 1}{\sigma_U}} \right]^{\frac{\sigma_U}{\sigma_U - 1}}, \tag{7}$$

where C_{usmt} is the consumption of UPC u from store s at time t; φ_{usmt} is the quality of UPC u at store s at time t; U_{ismt} is the set of UPCs within food item i at store s at time t; and σ_U is the constant elasticity of substitution across UPCs within food item i in the store.

We normalize the quality given that the utility is homogeneous of degree one in quality. Following the literature, we normalize it as follows:

$$\left(\Pi_{u \in U_{ismt}} \varphi_{usmt} \right)^{\frac{1}{N_{ismt}}} = 1 \tag{8}$$

$$\left(\Pi_{i \in I_{smt}} \varphi_{ismt} \right)^{\frac{1}{N_{smt}}} = 1, \tag{9}$$

where N_{ismt} is the number of barcodes in food item i in store s at time t, and N_{smt} is the number of food items sold in store s at time t. Thus, we normalize the geometric mean of barcode quality as well as the geometric mean of item quality to be equal to one for each store and period.

With the utility function defined, we now proceed to address the lowest-tier problem: allocating expenditure across UPCs within a given food item, store, and MSA.

5.1.2 Allocating Expenditure across UPCs within Food Items

In the lowest tier of demand, the representative consumer allocates expenditure across barcodes within a given food category in a given retailer. Barcode u has the sales share S_{usmt} in item i at store s at time t as follows:

$$S_{usmt} = \frac{(P_{usmt}/\varphi_{usmt})^{1-\sigma_U}}{\sum_{k \in U_{ismt}} (P_{kist}/\varphi_{kist})^{1-\sigma_U}}, \tag{10}$$

where P_{usmt} is the price and φ_{usmt} is the quality of UPC u at store s at time t.

The corresponding price index for food item at store s at time t is as follows:

$$P_{ismt} = \left[\sum_{k \in U_{ismt}} \left(\frac{P_{ksmt}}{\rho_{ksmt}} \right)^{1-\sigma_U} \right]^{\frac{1}{1-\sigma_U}}. \tag{11}$$

5.1.3 Allocating Expenditure across Food Items within Stores

Next, we allocate expenditure across food items in a given store. The sales share of food item i in store s at time t is given by:

$$S_{ismt} = \frac{(P_{ismt}/\varphi_{ismt})^{1-\sigma_I}}{\sum_{k \in I_{smt}} (P_{ismt}/\varphi_{ismt})^{1-\sigma_I}}, \tag{12}$$

where P_{ismt} is the price and φ_{ismt} is the food item i sold by store s at time t.

Again, the corresponding price index for store s at time t is as follows:

$$P_{smt} = \left[\sum_{k \in G_{smt}} \left(\frac{P_{ksmt}}{\rho_{ksmt}} \right)^{1-\sigma_I} \right]^{\frac{1}{1-\sigma_I}}. \tag{13}$$

5.1.4 Allocating Expenditure across Stores within an MSA

Lastly, we solve the allocation of expenditure across stores within a given MSA. The sales share of store s within an MSA at time t is given by:

$$S_{smt} = \frac{(P_{smt}/\varphi_{smt})^{1-\sigma_S}}{\sum_{k \in S_{mt}} (P_{kt}/\varphi_{kt})^{1-\sigma_S}}, \tag{14}$$

where P_{smt} is the price and φ_{smt} is quality of store s at time t.

Again, the corresponding price index for store s at time t is as follows:

$$P_{mt} = \left[\sum_{k \in S_{mt}} \left(\frac{P_{kmt}}{\rho_{kmt}} \right)^{1-\sigma_S} \right]^{\frac{1}{1-\sigma_S}}. \tag{15}$$

5.1.5 Barcode Demand

Lastly, defining the sales for barcode u at store s in MSA m at time t as E_{usmt} and the retail sales in MSA m at time t as E_{mt}, it is given by:

$$E_{usmt} = S_{usmt} S_{ismt} S_{smt} E_{mt}. \tag{16}$$

Then, the quantities sold for barcode u can be written as

$$Q_{usmt} = \frac{E_{usmt}}{P_{usmt}} \tag{17}$$

Rephrasing it with (10), (12), (14), and (16), we have the following:

$$Q_{usmt} = \varphi_{usmt}^{\sigma_U - 1} \varphi_{ismt}^{\sigma_I - 1} \varphi_{smt}^{\sigma_s - 1} P_{usmt}^{-\sigma_U} P_{ismt}^{\sigma_U - \sigma_I} P_{smt}^{\sigma_I - \sigma_s} P_{smt}^{\sigma_s - 1} E_{mt}. \tag{18}$$

5.1.6 Retailer Problem

Let's define the retail chain as the parent company that owns local stores and suppose that they decide optimal prices in each unit of stores by taking into account substitutability acorss all the stores it owns. Furthermore, let's allow them to be large enough to internalize their effects on the MSA price index, but small relative to the overall MSA economy to take the MSA-level expenditure and factor prices as given. Note that the internalization of the impact on the MSA price index depends on the retailers' market share, despite assuming CES demand. Therefore, this makes the retail chains face the elasticities of demand varying across the chain market share.

Let V_{usmt} denote the total variable cost for supplying barcode u in store s. Then it follows:

$$V_{usmt}(Q_{usmt}) = z_{usmt} Q_{usmt}^{1+\delta_i}, \tag{19}$$

where Q_{usmt} is the total quantity of barcode u in store s at t; δ_g determines the convexity of marginal cost with respect to output for barcodes in product item i; and z_{usmt} is a store-barcode-specific cost shifter. Costs are incurred in terms of a composite factor input that is assumed as a numeraire. This structure is consistent with Broda and Weinstein (2010), Burstein and Hellwig (2007), and Hottman (2017).

Suppose that each retailer store in MSA m needs to pay a fixed operating cost of F_{mt}. The profit of retail chain r in MSA m at time t is as follows:

$$\pi_{rmt} = \sum_{u \in U_{rmt}} P_{urmt} Q_{urmt} - V_{urmt}(Q_{urmt}) - F_{mt}, \tag{20}$$

where U_{rmt} is the set of barcodes sold in MSA m at time t at stores owned by retail chain r.

In the case of Bertrand competition, each retail chain chooses their prices $\{P_{urmt}\}$ to maximize profits. The first-order conditions take the following form:

$$Q_{usmt} + \sum_{k \in U_{rmt}} \left(P_{ksmt} \frac{\partial Q_{ksmt}}{\partial P_{ksmt}} - \frac{\partial V_{ksmt}(Q_{ksmt})}{\partial Q_{ksmt}} \frac{\partial Q_{ksmt}}{\partial P_{ksmt}} \right) = 0. \tag{21}$$

The optimal price is then given by

$$P_{usmt} = \mu_{rmt} m_{usmt}, \tag{22}$$

where μ_{rmt} is a markup, which is common across all products within retail chain r in MSA m at time t, over the marginal cost m_{usmt} of selling UPC u in store s in time t as follows:

$$m_{usmt} = z_{usmt}(1 + \delta_i) Q_{usmt}^{\delta_i}.$$

This markup is characterized by

$$\mu_{rmt} = \frac{\epsilon_{rmt}}{\epsilon_{rmt} - 1}, \tag{23}$$

where ϵ_{rmt} is the perceived elasticity of demand for retailer r in MSA m at time t. This is given by

$$\epsilon_{rmt} = \sigma_s - (\sigma_s - 1) S_{rmt}, \tag{24}$$

where σ_s is the constant elasticity of substitution across stores in MSA m and S_{rmt} is the market share of retail chain r in MSA m in time t.[18] Therefore, retailers with a higher sales share have higher markups, allowing them to increase prices more than retailers in less concentrated markets.

5.2 Alternative Hypothesis

An alternative hypothesis for spatial disparities in inflation revolves around cost differentials. If marginal costs in lower-income areas rise faster than in higher-income areas, this can contribute to higher inflation in those regions, irrespective of the market structure of retailers. To test this hypothesis, we use wage data for retail workers from the American Community Survey (ACS) and compare wage level and growth across MSAs with varying income levels.

We estimate the following two regressions to examine wage variations and growth across MSAs with different income levels:

$$\ln w_{mt} = \beta_0 + \beta_1 Income_{mt} + \delta_m + X'_{mt}\gamma + \delta_t + \varepsilon_{mt}$$

$$\Delta \ln w_{mt} = \beta_0 + \beta_1 Income_{mt} + \delta_m + X'_{mt}\gamma + \delta_t + \varepsilon_{mt},$$

where $\ln w_{mt}$ ($\Delta \ln w_{mt}$) is log wage (or the growth of wage) in retail sector in MSA m and year t, $Income_{mt}$ is the income per capita of MSA m in year t, and X_{mt} is a vector of MSA-level characteristics, including the share of college workers in retailers, the number of large retailers (with 500 or more employees), and δ_m and δ_t is the MSA and year fixed effects, respectively.

Table 12 displays the results. The top panel shows the results for the average wage, and the bottom panel presents the findings for wage growth. These results reveal that the average wage is generally lower in lower-income areas, even after controlling for the composition of skills and the share of large firms. However, the second panel suggests there are no significant patterns in wage growth across MSAs. While the data remains aggregated, it provides suggestive evidence that retailer wages are neither higher nor growing faster in lower-income areas, which helps rule out the cost-related channel to account for inflation heterogeneity.

[18] Note that if assuming Cournot competition, the elasticity of substitution ϵ_{rmt} becomes $\epsilon_{rmt} = \frac{1}{\frac{1}{\sigma_s} - (\frac{1}{\sigma_s} - 1)S_{rmt}}$, and if the sales share of retail chain approaches zero, the markup becomes the standard CES markup of $\frac{\sigma_s}{\sigma_s - 1}$.

Table 12: Average Wage and Growth in Retail Sector across MSAs

	Retail Wage	Retail Wage	Retail Wage
Income	0.400***	0.395***	0.413***
	(0.078)	(0.075)	(0.075)
College Share		0.711***	0.713***
		(0.048)	(0.048)
Large Firm Share			-15.52***
			(5.747)
Constant	8.934***	8.837***	8.775***
	(0.284)	(0.273)	(0.273)
MSA FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,868	2,868	2,868
	ΔRetail Wage	ΔRetail Wage	ΔRetail Wage
Income	- 0.032	-0.037	-0.026
	(0.123)	(0.121)	(0.121)
College Share		0.679***	0.681***
		(0.074)	(0.074)
Large Firm Share			-11.64
			(10.14)
Constant	0.128	0.033	-0.004
	(0.451)	(0.443)	(0.444)
MSA FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,580	2,580	2,580

Note: The table presents MSA-level wage regression results. The dependent variable in the top panel is the average log wage, while the bottom panel shows the log difference in wages within the retail sector. The sample period spans from 2006 to 2016, with MSA and quarter fixed effects included. In the second and third columns, the share of college-educated workers is included as a control, and in the third column, the share of large firms (those with 500 employees or more) is additionally controlled. *** p<0.01, ** p<0.05, * p<0.1

6 Concluding Remarks

In this paper, we investigate spatial variation in inflation and the role of retailer market structure in explaining these disparities. Using data from the NielsenIQ Retail Scanner and the Business Dynamic Statistics, we find that poorer MSAs experienced higher food inflation than wealthier areas. These regions are also characterized by fewer goods, fewer retailers, and greater market concentration. Using a difference-in-differences identification strategy, we provide causal evidence that retailer concentration contributes to regional inflation disparities. This suggests that market concentration plays a key role when retailers face cost shocks, allowing them to pass on higher prices to consumers. Additionally, through a model incorporating nested CES preferences and Bertrand competition, we demonstrate that higher retailer market power, as a result of increased market concentration, is a likely driver of these inflationary trends through higher markups. Our findings have important policy implications for real income inequality and highlight the limitations of official inflation indexes.

This paper serves as the initial step in a broader research agenda. Moving forward, we plan to further investigate the relationship between the inflation gap across MSAs and retailer market power by directly estimating markups from the data. Additionally, we aim to expand our analysis to explore the role of labor market power within the retail sector, examining how the interplay between product and labor market power can influence real income inequality and affect consumer welfare. These areas will be the focus of our future research.

References

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

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 02 2020, *135* (2), 645–709.

Broda, Christian and David E Weinstein, "Product creation and destruction: Evidence and price implications," *American Economic Review*, 2010, *100* (3), 691–723.

Burstein, Ariel and Christian Hellwig, "Prices and market shares in a menu cost model," 2007.

Cao, Yue, Judith A Chevalier, Jessie Handbury, Hayden Parsley, and Kevin R Williams, "Distributional Impacts of the Changing Retail Landscape," Working Paper 33307, National Bureau of Economic Research December 2024.

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

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

Haltiwanger, John, "Job Creation and Firm Dynamics in the United States," *Innovation Policy and the Economy*, 2012, *12*, 17–38.

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

Hobijn, Bart and David Lagakos, "Inflation inequality in the United States," *review of income and Wealth*, 2005, *51* (4), 581–606.

_, Kristin Mayer, Carter Stennis, and Giorgio Topa, "Household inflation experiences in the US: a comprehensive approach," in "in" Federal Reserve Bank of San Francisco 2009.

Hottman, Colin J., "Retail Markups, Misallocation, and Store Variety Across U.S. Cities," in "in" 2017.

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *Quarterly Journal of Economics*, 12 2018, *134* (2), 715–783.

Kaplan, Greg and Guido Menzio, "The morphology of price dispersion," *International Economic*

Review, 2015, *56* (4), 1165–1206.

— **and Sam Schulhofer-Wohl**, "Inflation at the Household Level," *Journal of Monetary Economics*, 2017, *91*, 19–38. The Swiss National Bank/Study Center Gerzensee Special Issue: "Modern Macroeconomics: Study Center Gerzensee Conference in Honor of Robert G. King" Sponsored by the Swiss National Bank and the Study Center Gerzensee.

Kerns-D'Amore, Joey Marshall Kristin and Brian McKenzie, "United States Migration Continued to Decline from 2020 to 2021," 2022. Accessed: 2024-10-30.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Martin, Robert S., "Democratic Aggregation: Issues and Implications for Consumer Price Indexes," *Review of Income and Wealth*, 2024, *n/a* (n/a).

Molloy, Raven, "JUE insight: Differences in rent growth by income from 1985 to 2021 and implications for inflation," *Journal of Urban Economics*, 2024, *142*, 103669.

Moretti, Enrico, "Real Wage Inequality," *American Economic Journal: Applied Economics*, January 2013, *5* (1), 65–103.

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

Schanzenbach, Diane Whitmore, Ryan Nunn, Lauren Bauer, and Megan Mumford, "Where does all the money go: Shifts in household spending over the past 30 years," *Brookings Institution, The Hamilton Project*, 2016, *31.*

Smith, Dominic A and Sergio Ocampo, "The evolution of US retail concentration," *American Economic Journal: Macroeconomics*, 2025, *17* (1), 71–101.

Stroebel, Johannes and Joseph Vavra, "House prices, local demand, and retail prices," *Journal of Political Economy*, 2019, *127* (3), 1391–1436.

Appendix

A Robustness with Nielsen Consumer Panel

Table A.1: Fraction of Households Shopping Outside of their Residential MSAs

Indicator	Observation	Percent
1	780,500	92.32
0	64,932	7.68
Total	845,432	100

Notes: The table shows the fraction of households that consume outside their residential MSAs (with an indicator value of 1) in each given year. The data covers household-year observations from 2006 to 2020.

In this section, we use the Consumer Panel data, which contains individual-level demographic and purchase information from Nielsen. The analysis utilizes the household-year level sample from 2006 to 2020 and identifies households that make purchases outside their residential MSAs in a given year. The results in Table A.1 show that, on average, 92% of households made purchases exclusively within their residential MSAs.

Furthermore, when examining household characteristics and shopping patterns by each category, Table A.2 shows that their properties (such as income levels, the average number of stores households purchase from, and total amount of spending) are similar across groups. For households that shop outside of their MSAs, they visit an average of 1.75 stores, spend approximately 50% of their total expenditure outside their residential MSAs, and the average number of these outside MSAs they purchase from is 1.05.

In addition, we compute income deciles using two different MSA definitions in Nielsen. One is based on the MSA information of households in the Consumer Panel, and the other is based on the MSA information of consumers, derived by linking the locations of stores from which households make purchases in the Scanner data with household income data in the Consumer Panel. Table A.3 shows the gap between these two definitions, revealing that most MSAs (75.27%) align with the same income decile definitions, and only a very small fraction (0.54%) exhibit a gap of three deciles. This confirms that our baseline measures of income deciles based on store locations and

Table A.2: Characteristics of Households by Shopping Types

Variable	Mean (Std.)	Mean (Std.)
Indicator	0	1
Income	20.46	19.94
	(5.98)	(5.87)
Store #	3.32	3.77
	(1.90)	(2.08)
Spending Amount	1812.05	1659.12
	(1985.68)	(1811.08)
Store # (out)		3.77
		(2.08)
Spending Amount (out)		714.78
		(1226.29)
MSA # (out)		1.05
		(0.23)
Obs	780,500	64,932

Notes: The table provides the shopping characteristics of households by their types based on whether they shop outside of their residential MSAs (indicator=1) or not (indicator=0). The second column indicates the households only shopping inside their MSAs, and the last column shows those shopping outside of their MSAs. Store # is the number of stores the households purchase from, Spending Amount is the total amount of spending, Store # (out) is the number of stores outside of the household's living MSAs, Spending Amount (out) is the amount of spending made outside of their living MSAs, and MSA # (out) is the number of MSAs the shop, outside of their living MSAs. This is the household-year level sample over 2006-2020.

Table A.3: Gaps in Two Income Decile Definitions: Household vs. Consumer MSAs

Gap	Observation	Percent
-3	1	0.54
-1	20	10.75
0	140	75.27
1	25	13.44
Total	186	100

Notes: The table computes the gap in income deciles when defined by consumer income and household income, using an MSA-level sample.

BEA income per capita data are not mismeasured.

B Robustness for Other Food Items and Indexes

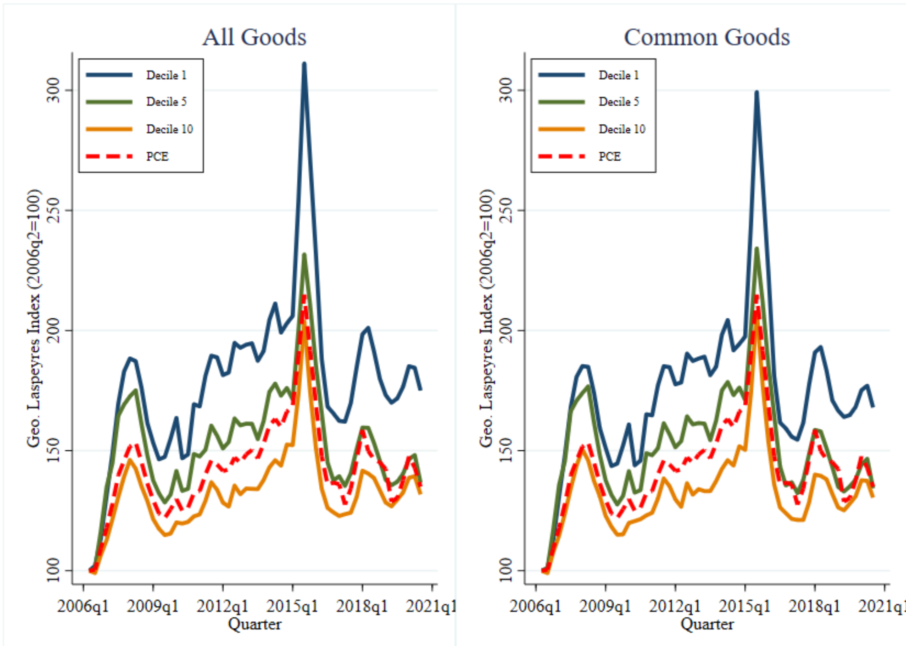


Figure B.1: Laspeyres Price Index for Eggs

Notes: The figure represents relative prices in the aggregated egg market with five series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2020Q4. The data come from NielsenIQ Retail Scanner dataset represented by geometric Laspeyres price indexes. Each solid line corresponds to a decile of the income per capita ranking of MSAs with decile 1 containing the MSAs with the lowest income per capita and decile 10 containing the MSAs with the highest income per capita. The red dashed line corresponds to the official PCE price index from the U.S. Bureau of Economic Analysis. The left panel is the set of goods sold at retailers in quarters t and t-1. The right panel corresponds to the set of goods present across all 10 deciles in quarters t and t-1. We map the NielsenIQ UPCs to the PCE definition of eggs by using a product module concordance provided by the BLS.

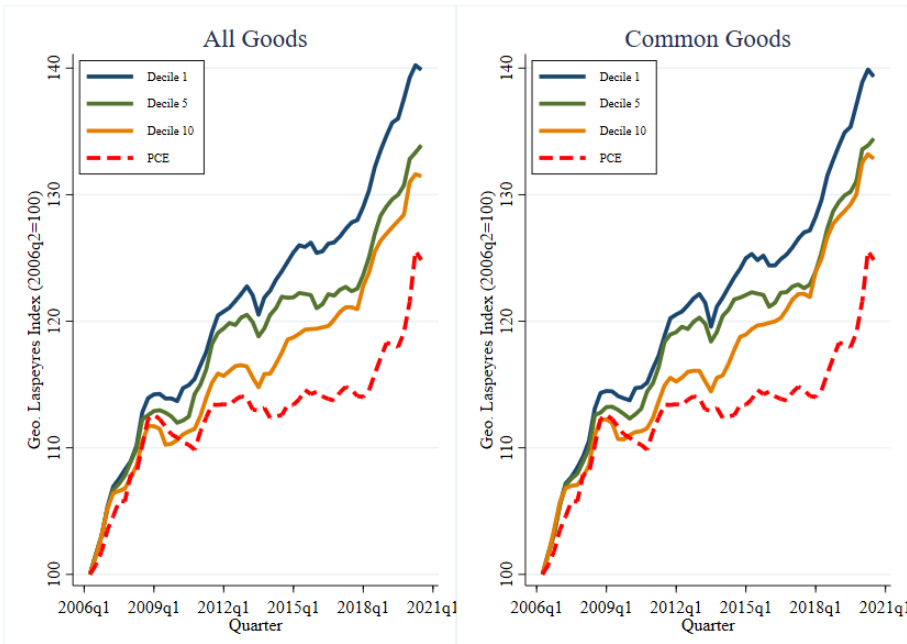
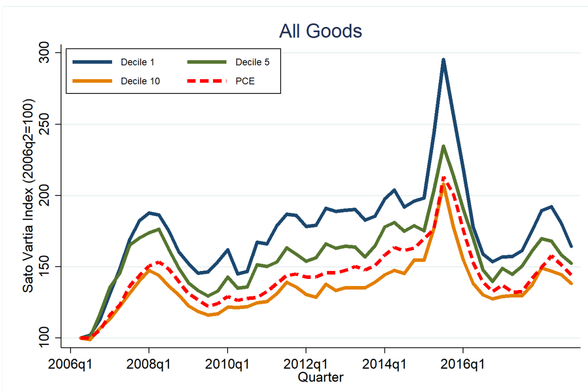
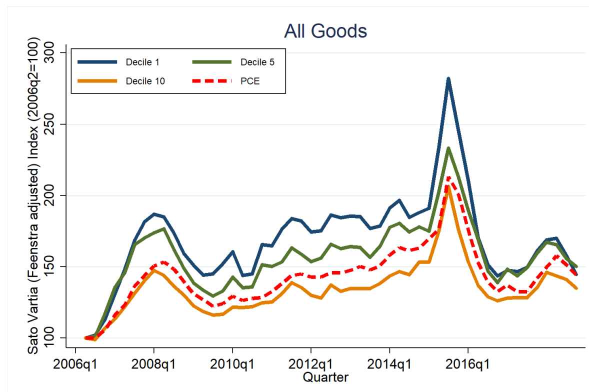


Figure B.2: Laspeyres Price Index for Soda and Juices

Notes: The figure represents relative prices for the soda and juices market with five series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2020Q4. The data come from NielsenIQ Retail Scanner dataset represented by geometric Laspeyres price indexes. Each solid line corresponds to a decile of the income per capita ranking of MSAs with decile 1 containing the MSAs with the lowest income per capita and decile 10 containing the MSAs with the highest income per capita. The red dashed line corresponds to the official PCE price index from the U.S. Bureau of Economic Analysis. The left panel is the set of goods sold at retailers in quarters t and t-1. The right panel corresponds to the set of goods present across all 10 deciles in quarters t and t-1. We map the NielsenIQ UPCs to the PCE definition of soda and juices by using a product module concordance provided by the BLS.



Sato Vartia



Feenstra-adjusted Sato Vartia

Figure B.3: Demand-based Price Indexes for Eggs

Notes: The figure represents relative prices for the aggregated egg market with five series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2016Q4. The data come from NielsenIQ Retail Scanner dataset represented by Sato-Vartia and Feenstra-adjusted Sato-Vartia price indexes. Each solid line corresponds to a decile of the income per capita ranking of MSAs with decile 1 containing the MSAs with the lowest income per capita and decile 10 containing the MSAs with the highest income per capita. The red dashed line corresponds to the official PCE price index from the U.S. Bureau of Economic Analysis. We map the NielsenIQ UPCs to the PCE definition of eggs by using a product module concordance provided by the BLS.

C Robustness with BDS

Table C.1: The Share of Large Firms (BDS)

	Large firm share	Large firm emp. share
Income	-0.044***	-0.049***
	(0.006)	(0.009)
Year FE	Yes	Yes
MSA FE	Yes	Yes
Observations	8,001	8,001

Notes: The table represents regression results from our two-way fixed effects estimator. The coefficient of interest is the coefficient on income per capita (in $1000) in an MSA for a given year. The dependent variable is an unweighted share (%) of large retailers in column 1 and an employment weighted large share (%) of large retailers in column 2. Data is collected from the Business Dynamics Statistics and retailers are gathered from retail trade sector (NAICS 44-45) for 2000-2020. *** p<0.01, ** p<0.05, * p<0.1

We run the following regression using the BDS dataset to confirm the cross-sectional pattern:

$$LargeFirm_{mt} = \beta_0 + \beta_1 Income_{mt} + \delta_m + \delta_t + \varepsilon_{mt},$$

where $LargeFirm_{mt}$ is the (employment) share of large firms in MSA m in year t. $Income_{mt}$ is income per capita in MSA m, and δ_m and δ_t are MSA and year fixed effects, respectively. The results, displayed in Table C.1, show that larger firms are more prevalent and have a higher share of employment within lower income deciles.

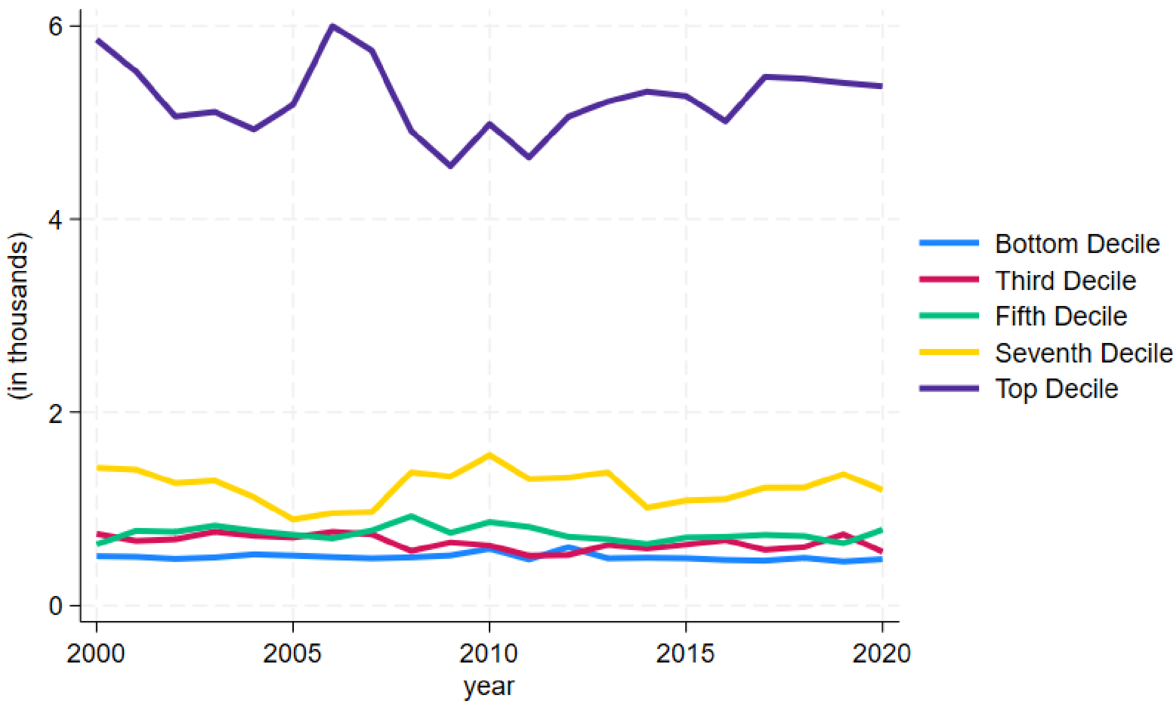


Figure C.1: Number of Retail Chains by Income Decile

Notes: The figure represents the number of retail chains present from 2000 to 2020 in deciles 1, 3, 5, 7, and 10. The data on the number of retail chains come from the Business Dynamics Statistics (BDS). We specifically only use data on chains from the BDS for the retail trade sector (NAICS codes 44-45).

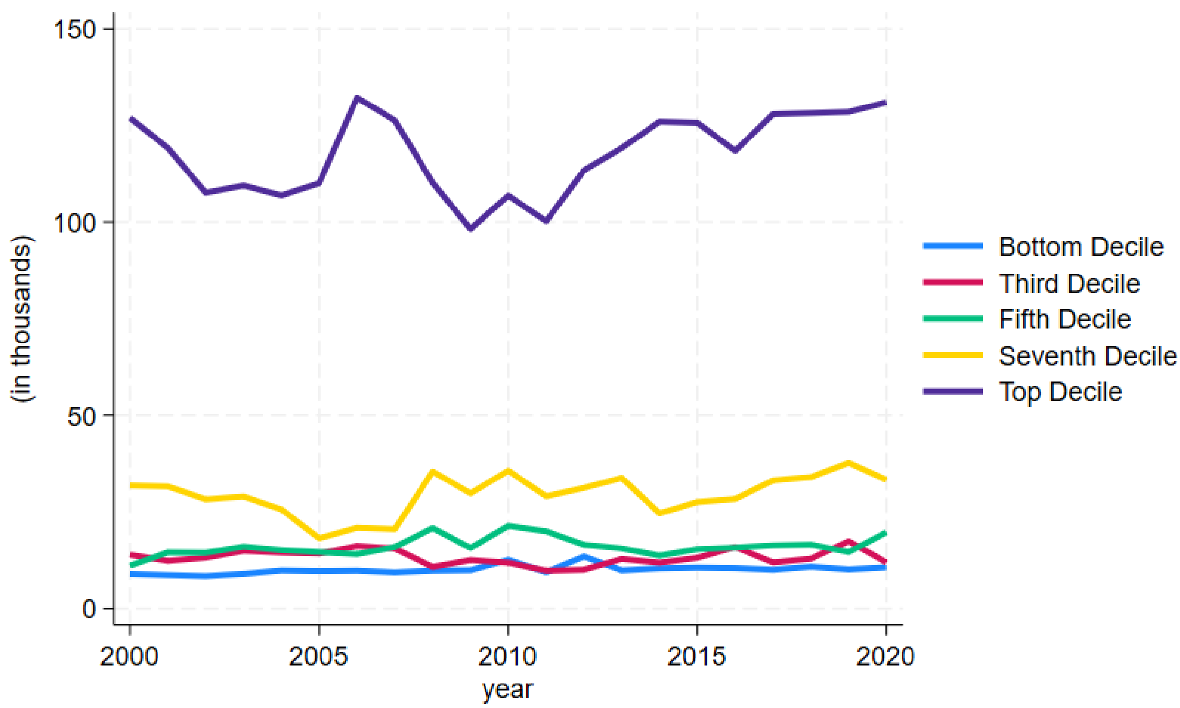


Figure C.2: Employment of Retail Chains by Income Decile

Notes: The figure represents the employment of retail chains present from 2000 to 2020 in deciles 1, 3, 5, 7, and 10. The data on the number of retail chains come from the Business Dynamics Statistics (BDS). We specifically only use data on chains from the BDS for the retail trade sector (NAICS codes 44-45).

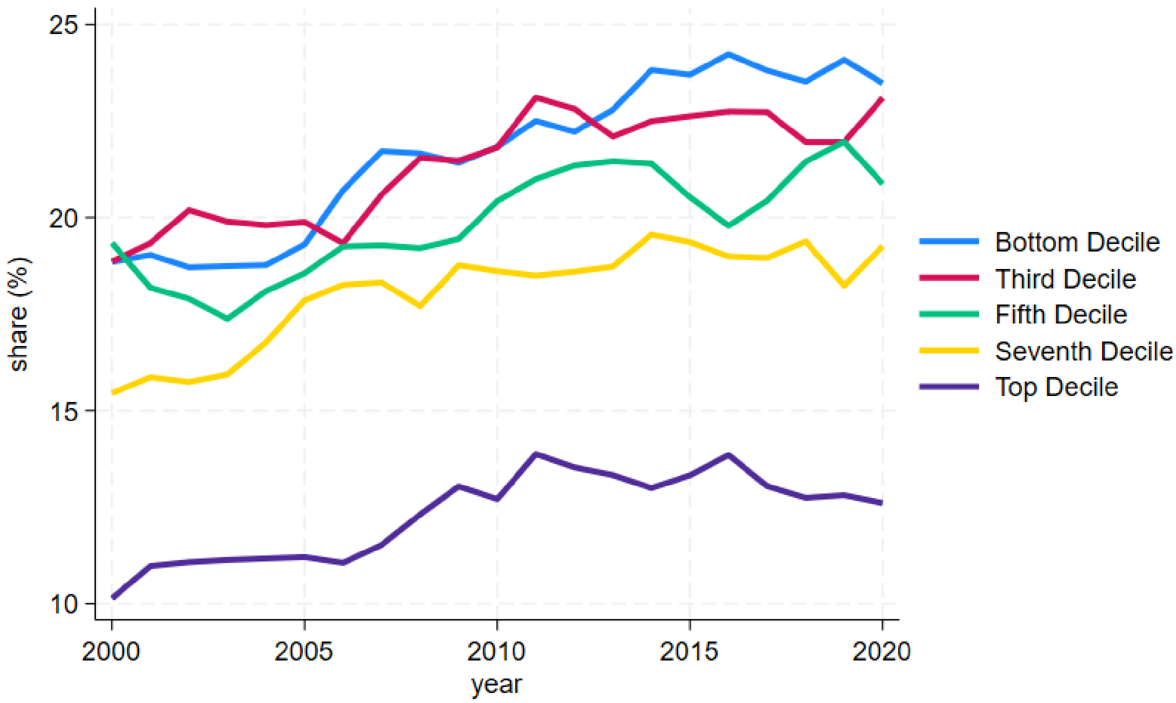


Figure C.3: Share of Large Retailers

Notes: The figure represents the share of establishments from large retail chains present from 2006 to 2019 in deciles 1, 3, 5, 7, and 10. A large retail chain is defined as having (at the firm level) more than 500 employees. The data on the number of retail chains come from the Business Dynamics Statistics (BDS). Note that the firm-level definition of large retailers is based on the number of employees at the firm level (national), while the share of large retailers is defined using information at the establishment level (MSA), specifically, only using data on chains from the BDS for the retail trade sector (NAICS codes 44-45).

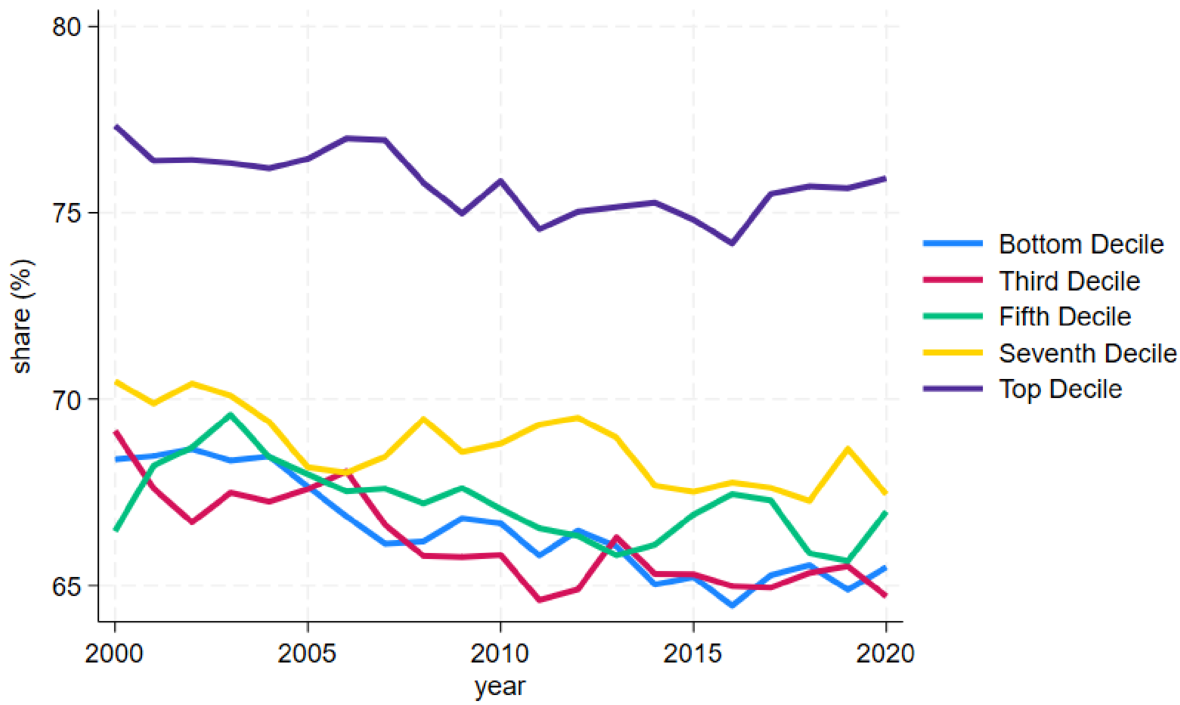


Figure C.4: Share of Small Retailers

Notes: The figure represents the share of establishments from small retail chains present from 2006 to 2019 in deciles 1, 3, 5, 7, and 10. A small retail chain is defined as a retail chain (firm level) that has fewer than 20 employees. The data on the number of retail chains come from the Business Dynamics Statistics (BDS). Note that the firm-level definition of small retailers is based on the number of employees at the firm level (national), while the share of small retailers uses information at the establishment level (MSA). We specifically only use data on chains from the BDS for the retail trade sector (NAICS codes 44-45).