Geospatial Heterogeneity in Inflation: A Market Concentration Story*

Seula Kim†
Penn State University and IZA

Michael A. Navarrete‡
NBER

December 9, 2024

[Click for the Latest Version]

*This paper was previously circulated under the title "Geospatial Heterogeneity in Inflation: The Role of Retailer Dynamics." We thank Miguel Ampudia (discussant), Ben Bernanke, Marcus Casey, Thesia Garner, John Haltiwanger, Judith Hellerstein, Colin Hottman, David Johnson, Ethan Kaplan, Munseob Lee, Giacomo Ponzetto, Lumi Stevens, and participants at seminars at BLS, the University of Maryland, the BdE-BOC-ECB-Fed Conference, the Washington, DC Urban Economics Workshop, the NASMES meeting, the BSE Summer Forum, UCL Stone Centre PhD Conference, the UChicago-SNU Inequality conference, the Federal Reserve Board, and the Midwest Macro meeting for helpful comments. We also thank the RESET team, without whom this project would not have been feasible, 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; Website: https://www.seulakim.com

‡Email: navarrem@nber.org; Website: https://mnav-umd.github.io

Abstract

This paper studies how spatial variation in inflation rates affects real income inequality and examines the role of retailer dynamics in driving these differences. Using the NielsenIQ Retail Scanner dataset and the Business Dynamic Statistics, we document several stylized facts about the spatial heterogeneity in inflation and retailer dynamics. We find that poorer MSAs experienced higher inflation than richer MSAs on average from 2006 to 2020. The differences are substantial: the annualized difference in inflation between the poorest MSAs and the richest MSAs is 0.46 percentage points (10 p.p. in total over the period). Poorer MSAs have fewer retailers and less variety of goods with a larger fraction of large retailers; these poorer MSAs have higher retailer market concentration relative to richer MSAs. To explore a potential causal link between inflation and market concentration, we use a triple-difference estimator, with a particular focus on the egg market during the 2014-2015 bird flu episode. Our analysis suggests that retailer market concentration contributes to the difference in inflation between poor and rich MSAs.
JEL Code: E31, I31, J60

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

1 Introduction

Inflation is an important economic indicator that can have significant implications for economic growth and stability. However, both the literature on inflation and policymakers often overlook the heterogeneity in inflation rates across different subnational regions, in particular within disaggregated food categories. Our research seeks to address this gap by first documenting the heterogeneity in inflation rates across metropolitan statistical areas (MSAs) for each personal consumption expenditure (PCE) food item and then investigating the relationship between inflation and market concentration to understand a novel mechanism through which retailer dynamics affect inflation. One reason we study food is the potentially larger price dispersion in food prices than in other goods such as consumer technologies. Specifically, there may be more market segmentation in food because individuals typically have to travel to a local store rather than purchasing a product online.

Using the NielsenIQ Retail Scanner data, we make three primary contributions to the literature: (1) We document the new stylized fact that poorer MSAs face higher inflation in food; (2) we find suggestive evidence that this inflation gap across regions is partially due to retail market power; and (3) we document price dispersion within universal product codes (UPCs) across MSAs. We use NielsenIQ Retailer Scanner data to construct price indexes across MSAs, which allows us to use highly disaggregated data, 12-digit universal product codes. UPCs are highly disaggregated such that two cans of Campbell's tomato soup in different sizes would be associated with two different UPCs. We use a quasi-experiment of the 2014-2015 bird flu episode to directly link market concentration to inflation. Even though we find higher inflation in poorer MSAs, we find lower price levels in the poorer MSAs. These results are at odds with some previous work (DellaVigna and Gentzkow, 2019), which we are able to reconcile. The regional variation in inflation we identify implies that real income inequality, assuming uniform inflation across the U.S., will understate the disparity when heterogeneous inflation across areas is considered. Specifically, the real income gap between the top and bottom deciles will widen if we use regional price deflators and 2006 real income as the baseline.

Our findings indicate that food inflation rates vary across regions with different income levels. In particular, on average, the poorest decile of MSAs exhibit higher inflation rates than richer MSAs over the period from 2006 to 2020.[1] The cumulative difference in inflation between the bottom decile and the top decile over this period is about 10 percentage points. Furthermore, we show that within poorer areas, the fraction of large retailers (with 500 or more employees) is higher and the fraction of small retailers (with 19 or fewer employees) is lower. The opposite pattern is observed within richer areas. Relatedly, we find that retailer sales are more concentrated in poorer areas. Market concentration is also associated with higher inflation rates within an MSA-PCE disaggregated food category. Generally, this gap in inflation rates increases the larger the income per capita gap between MSAs being compared. In other words, the MSAs in the bottom decile in terms of income per capita are the MSAs with the largest price changes over our sample period and are also the MSAs with the highest degree of market concentration.

These patterns are robust to disaggregated and aggregated food items both, as well as to imposing a common goods rule that all 10 deciles of MSAs have the same universal product codes (UPCs) in their consumption basket. This restriction partially reduces the differences between the poor and rich MSAs. In general, imposing the common goods rule restricts the UPCs in the richest income deciles but has little effect on the basket of goods available in poorer income deciles. Typically the UPCs available in the poorer deciles are a subset of the UPCs available in the richer deciles.

Although not required for heterogeneous inflation rates across regions, we find price dispersion at the UPC level across regions. Even though poorer regions experience relatively higher inflation rates than richer areas, we find that prices in poorer regions are relatively lower than in richer MSAs. Our results partially contrast with the uniform

[1] Note that the largest differences in inflation rates between the top and bottom deciles is more pronounced prior to 2016.

pricing literature (DellaVigna and Gentzkow, 2019). Despite having an overall lower price level, higher inflation in poorer regions should still be a concern for real income inequality. Specifically, this convergence in prices is not followed by a convergence in incomes. Furthermore, individuals in these poorer regions are likely to be more sensitive to inflation in food prices than individuals in richer regions given that a higher share of their spending is allocated to food expenditures. For example, individuals in West Lafayette, Indiana would have to allocate a higher share of their income if food prices increase by 1% than individuals in Minneapolis, Minneapolis facing the same percentage increase. Not only do we find that individuals in poorer MSAs have fewer retailers to choose from, but these individuals also have access to fewer products (UPCs).

These findings have important implications for policymakers across multiple dimensions. First, understanding heterogeneity in inflation rates can inform monetary policy and consumer protections. Official government price indexes are aggregated to the national level, which can misrepresent inflation in the poorest regions of the country. Specifically, these price indexes are aggregated using expenditure weights, and the rich areas account for a disproportionate share of expenditures. Thus, relying on aggregate indexes may underrepresent inflation experienced in the poorest regions and overrepresent the richest areas. By focusing exclusively on aggregate measures, policymakers could make policy decisions that are not reflective of most of the country (population weighted).

Furthermore, this spatial analysis highlights that aggregates can mask heterogeneity. The heterogeneity in inflation is important partly because it is costly for individuals to move. According to the U.S. Census Bureau report, the proportion of Americans moving has declined since the 1980s (Kristin Kerns-D'Amore and McKenzie, 2022). The concern is that higher inflation in food and beverages is hurting vulnerable Americans who are not mobile and have fewer products to substitute towards. Unlike other types of products, food and beverages are a necessity, which is why Americans may be particularly sensitive to such price increases. Further, all Americans consume and repeatedly purchase food and beverages. The combination of market segmentation and retail market power creates the potential for retailers to pass a higher burden of cost shocks to consumers in these poorer regions.

2 Literature Review

Previous work has shown that higher income households experience lower inflation rates than poorer households through differences in consumption baskets (Jaravel (2018); Kaplan and Schulhofer-Wohl (2017)). At the same time, poor and rich households are commingled by living in the same geographic region. Handbury (2021) documents that the welfare difference between rich and poor households could depend on the set of goods available in each region, which gets exacerbated in wealthy cities that have the largest amenities. However, we find that regional differences in consumption baskets are not always on the consumer side given that consumers in poorer MSAs have fewer UPCs to choose from. Thus, our research contributes to this line of studies by shedding light on a new dimension of consumption inequality outside of consumption baskets. We identify a new source of heterogeneity in inflation: regional. We find another mechanism other than a demand-based love of variety that can account for these differences: market concentration. The regional variation is particularly interesting given that market concentration is inherently regional.

In addition, our paper expands beyond previous studies exploring potential sources of market concentration at a local market level. Market concentration is a novel channel through which inflation rates could vary. Previous differences in inflation rates across groups have been attributed to differences in consumption baskets (Jaravel, 2018). By incorporating market power and concentration as sources explaining part of the variation in heterogeneous inflation rates, we are able to add to the markups literature (Hottman (2017), Autor et al. (2020), and De Loecker et al. (2020). Nevo (2001) shows that collusion is not necessary for firms to charge high-price cost margins in the cereal market, which is characterized by high concentration. Feenstra et al. (2022)

find that the profit share of firms has been increasing over time. The increase in profit shares and the rise in markups could amplify the difference in inflation rates across high- and low-income areas. We aim to analyze retailer dynamics in these segmented local markets to see their correlation with the inflation dispersion across regions.

Lastly, our study is in line with 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 in what goods are being measured and which areas are being considered.[2] Previous 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, product modules, in food.

3 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 lets us measure inflation rates across regions by using sales information across retailers for food products. The BDS dataset allows us to see if market concentration is driving the patterns we observe.

[2]We use a narrower set of goods but are broader in the areas considered.

3.1 NielsenIQ Retail Scanner

Our analysis is based on the RMS dataset provided by the Kilts Center at Chicago Booth. The data consist of weekly pricing, volume, and store merchandising conditions generated by more than 100 retail chains across all U.S. markets, which includes over 40,000 individual stores. Total sales in the NielsenIQ RMS sample are worth over $200 billion per year and represent 50% of total sales in grocery stores, 55% in drug stores, 32% in mass merchandisers, and 2% in convenience stores.

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).[3] These ELIs then map to PCE disaggregated categories.

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

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.[4] We further de-

[3] We were provided this concordance as part of the Re-Engineering Statistics using Economic Transactions (RESET) project.

[4] Note that our baseline analysis relies on the MSA location of retailer stores in NielsenIQ. There may be potential concerns with this measure if an MSA is broad enough to encompass consumers who move across MSAs, potentially creating a gap between the income of consumers and that of residents. To address this, we leverage the Consumer Panel data to examine the fraction of households shopping outside of their residential MSAs and explore their characteristics. Furthermore, 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 C, which all help address potential concerns.

fine income deciles by the cross-time average of the MSA-level income per capita.[5] We then aggregate the data to the monthly frequency using the National Retail Federation calendar and aggregate it up to the quarterly level. Next, using the concordance between the product modules and the PCE food categories, we identify the food sector in NielsenIQ. Lastly, to measure manufacturer market power and degree of competition, we merge the quarterly data with manufacturer identifiers by UPC codes.[6] The steps we follow are similar to those used by Hottman et al. (2016).

Our main analysis is at the 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 income decile and food category-quarter. Table 3 provides summary statistics for the main sample.

3.2 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.[7] In the BDS, we use retailers' information (based on NAICS code 44-45) and construct a set of business dynamics measures at the MSA level.

[5] See the examples of the income deciles in Table 2.

[6] The manufacturer identifiers are provided by GS1, the company in charge of allocating barcodes.

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

3.3 Main Measures

3.3.1 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 Lapeyres price index as follows:

$$\ln \Psi_{mt}^{G} = \sum_{k \in \mathbb{C}_{mt-1,mt}} w_{mkt} \ln \frac{p_{mkt}}{p_{mkt-1}}, \tag{3.1}$$

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

Although our default measure is the geometric Laspeyres index, we also use the geometric Paasche index, replacing the weights with the current expenditure shares ($w_{mkt} = s_{mkt}$). We also do a robustness test using alternative demand-based indexes based on the constant elasticity of substitution (CES) preference assumption due to potential substitution bias associated with the traditional indexes.[8] One is the Sato-Vartia index, where we replace the above weight with $w_{kt} = \frac{\frac{(s_{k,t}-s_{k,t-1})}{(\ln s_{k,t}-\ln s_{k,t-1})}}{\sum_{k\in\mathbb{C}_{t-1,t}} \frac{(s_{k,t}-s_{k,t-1})}{(\ln s_{k,t}-\ln s_{k,t-1})}}$, which considers the demand effect for common goods appearing between $(t-1)$ and t. Another index is the Feenstra-adjusted Sato-Vartia index, which further considers the effect of product entry and exit. It is constructed using the following formula:

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

where $\lambda_{t,t-1} = \frac{\sum_{k\in\mathbb{C}_{t-1,t}} p_{k,t}q_{k,t}}{\sum_{k\in\Omega_t} p_{k,t}q_{k,t}}$, $\lambda_{t-1,t} = \frac{\sum_{k\in\mathbb{C}_{t-1,t}} p_{k,t-1}q_{k,t-1}}{\sum_{k\in\Omega_{t-1}} p_{k,t-1}q_{k,t-1}}$.[9]

[8] The traditional indexes do not take into account demand effects that may be generated from consumers' substitution across differentiated goods.

[9] For east of notation, we drop the MSA notation from the Sato-Vartia and the Feenstra-adjusted Sato-Vartia indexes.

Lastly, we also construct the price indexes by restricting our sample to UPCs sold in all 10 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 constructed using only the set of common goods to determine whether the spatial dispersion of price levels and growth comes from the difference in consumption baskets.

3.3.2 Retailer Dynamics

For the Nielsen IQ data, we define large and small chains based on the size distribution of the number of stores. We use store and retailer codes and geographic information to identify stores, retailers, and ownership structures (i.e., which retailer owns which stores across different regions and time). Specifically, we count the number of stores owned by retailers at the national level to proxy for retailer size. We define large chains by the top decile of chains based on the number of stores and small chains as the bottom decile. We calculate the number and share of large and small chains located in each MSA.

Alternatively, using the BDS, we define large and small retailers by their number of employees. We define large retailers as those with 500 or more employees, and small retailers as those with 19 or fewer employees. We then construct the share and employment share of large and small firms within each MSA and compare across different regions (MSAs).

Lastly, we use the sales share of retailers and construct the Herfindahl-Hirschman index (HHI) to obtain the degree of retailers' market concentration in each region.

4 Spatial Heterogeneity in Inflation and Retailer Dynamics

4.1 Price and Inflation Patterns

Figure 1 presents the geometric Laspeyres index constructed from the NielsenIQ Scanner, along with the official PCE price index across the first (poorest), fifth, and 10th (richest) income deciles. We focus on aggregated food. The left panel shows the price index including all UPCs, and the right panel only includes common goods, i.e., the UPCs that are present across all deciles. We set the base quarter to 2006Q2.

The general trend captured by Figure 1 is that the poorest decile (“Decile 1") exhibits higher price growth than the richer deciles (“Decile 5" and “Decile 10"). This pattern holds even after we restrict the sample to the set of common goods when constructing the price indexes. These results imply that the dispersion of price growth across deciles is not necessarily driven by different consumption baskets or by different preferences among consumers in different regions. These findings are generally consistent for the 21 PCE food categories as well as for other aggregated food series. See Figure 2 for eggs and Figure 3 for soda and juices, both of which are PCE disaggregate food categories. Furthermore, the patterns stay robust even after using the demand-based price indexes, as shown in Figure 4.

Lastly, the official PCE series is closer to the series for the highest income decile than it is to any other decile. Specifically, the official PCE price index series is understating inflation for individuals living in the poorest areas. This discrepancy in inflation has macroeconomic implications. For example, if we assumed uniform wage growth across the United States, then official real wage growth over this period is higher than actual real wage growth in the poorest areas.

4.2 Retailer Dynamics

To examine retailer dynamics across different regions, we compute summary statistics for our main sample from NielsenIQ by income-per-capita decile (Table 4). The table shows that richer income areas have more retailers and stores and that retailers in these areas have higher sales. In addition, the share of large chains is higher but the share of small chains is lower in poorer income areas. Finally, poorer income areas face a higher degree of sales market concentration.

We conduct additional analyses to further explore these patterns. Figure 5 exhibits the distribution of retailer store numbers in NielsenIQ across income deciles. In line with the results in Table 4, the poorer deciles have the smallest mass of small retailers (with fewer stores) relative to the richer deciles.

We next run the following regressions to explore cross-sectional retailer dynamics 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 5, confirm the cross-sectional patterns that richer areas have higher sales, more retailers and stores, and a lower fraction of large retailers.

We also find this pattern consistently in the BDS data. Figure 6 shows that more retail chains are located in richer areas, and Figure 7 shows that these retailers create more jobs in those areas. Furthermore, we find a clear pattern between firm size and income decile. Figures 8 and 9 present the share of large and small retailers, respectively, within each income decile. These figures indicate that poorer income areas have a larger fraction of large retailers and richer areas have a smaller share of large retailers. The reverse pattern is observed for the fraction of small retailers.[10]

[10]Note that the size of retailers is measured by firm-level employment, and the share is calculated

We also 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 6, show that larger firms are more prevalent and have a higher share of employment within lower income deciles.

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 represent fixed effects for the PCE food category and time, respectively. The results in Table 7 show that the HHI is higher in lower income deciles, indicating that poorer deciles experience a higher degree of market concentration.

Relatedly, we construct two versions of the HHI: one for all goods sold in any decile and another for common goods sold in all 10 deciles. 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.The results broadly suggest that market concentration varies across markets, but within each market, concentration is more skewed toward larger firms in the poorest areas. In addition, higher income areas tend to have higher elasticity of substitution in most markets, meaning that consumers are more willing to substitute different goods

based on the number of firms that operate retail stores in each MSA. This analysis is robust to using the number of establishments. Also, note that these patterns are robust across the whole sample period.

into their baskets.

Lastly, we also tabulate the total number of UPCs sold, common goods sales as a fraction of total sales, and the quantity of common goods as a fraction of total UPCs. We perform this analysis for each food category across income deciles. Consistent with our expectations, we find that poorer areas have fewer UPCs and have higher quantity and expenditure shares of total consumption allocated to the set of common goods.

4.3 Price Dispersion

Thus far, we have shown descriptive statistics indicating that poorer MSAs experience higher inflation as well as differences in retailer dynamics. However, variation in inflation across regions does not necessarily imply differences in price levels across areas. The differences in inflation rates could be entirely driven by differences in basket composition across areas. Unlike previous work (DellaVigna and Gentzkow, 2019), we find price dispersion in the eggs market, with poorer MSAs seeing lower prices.

In Figure 10, we look at the distribution of prices for the UPCs of eggs across stores in 2013q4. Panel (a) uses the entire sample, and Panel (b) is restricted to stores in the richest (San Francisco, CA) and poorest (McAllen, TX) MSA. Clearly, the distribution of San Francisco is to the right of McAllen's distribution. The gap between these MSAs persists even if we restrict our sample to the UPCs that are common to both MSAs. However, this restriction drastically limits the number of UPCs from San Francisco. Figure 10b provides suggestive evidence that poorer MSAs see lower prices for a given UPC.

The price dispersion between San Francisco and McAllen may not be systematically true across MSAs. To further investigate whether poorer MSAs are seeing lower prices, we examine all the UPCs of eggs across all stores from 2006q1 to 2020q4. In Table 9, we regress price levels on the average income per capita of a given MSA. In column 1, we control for UPC and quarter fixed effects and find that a $10,000 increase in the average income per capita of an MSA is a associated with a 1 cent increase in the price of a UPC. Note that prices are denominated on a per egg basis. In the second column, we impose a retailer fixed effect and see the association fall from 1 cent to 0.6 cent.

Even though we see lower prices for a given UPC in poorer MSAs, previous work has found uniform pricing, which could be due to restricting comparisons to the same retailers across areas. One issue with that comparison is that several of the retailers in the poorest MSAs are not present in the richest MSAs, which prevents comparisons from being made. One potential reason why retailers may charge different prices for a given UPC in different areas is due to market power and heterogeneous pass-through rates. Potentially, retailers in poorer MSAs are able to pass-through cost shocks to consumers at higher rates, allowing for price discrimination (Weyl and Fabinger, 2013).

5 Potential Mechanism through Retailer Market Concentration

To investigate a potential mechanism behind relatively higher inflation in poorer MSAs than richer MSAs, we perform additional analyses using HHI as our measure of market concentration. First, we look at the relationship between inflation rates and HHI. Then, to determine a causal relationship between the two, we exploit a quasi-experiment using a triple-difference estimator.

5.1 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:

$$P_{mt} = \beta_0 + \beta_1 HHI_{mt} + \delta_m + \delta_t + \varepsilon_{mt}, \tag{5.2}$$

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, respectively.

To measure whether there is an association between market concentration and inflation, we use the standard OLS estimator described in Equation (5.2) and present the results in Table 10 for the egg market. In column 1, when we control for MSA and quarter fixed effects, we find a positive relationship between HHI and inflation. This effect is significant at the 1% level.[11] However, we cannot speak to any causal relationships here as this analysis may contain an endogeneity bias. For example, this could be entirely demand driven where consumers in MSAs with higher HHI values could potentially prefer to consume goods that are experiencing relatively higher inflation.

We show the stylized fact that the poorer MSAs experienced higher inflation in food and beverages than the richer MSAs. However, we are not able to conclusively show what is driving this difference in inflation rates. One potential explanation is a supply-side story where poorer MSAs have fewer stores, which weakens competition and allows retailers to increase prices. A potential alternative explanation is a demand story where even after we restrict the analysis to the same set of goods across MSAs, the consumers in rich MSAs are different from consumers in poor MSAs. For example, consumers living in MSAs in the top decile could be more sensitive to price changes, leading firms to increase prices at slower rates.

To isolate whether the effect we find is coming from the supply side or demand side, we use the 2014–2105 bird flu outbreak as a quasi-experiment with a triple-difference estimator in the next section.

5.2 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 2 around 2014Q4–2015Q1. Based on U.S. Department of Agriculture (USDA) reports, 36 million layers (birds that lay eggs)

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

were lost due to the bird flu by June 2015.[12]

Importantly, USDA and the Government Accountability Office (GAO) reports state that the impact of the bird flu shock exhibits geospatial variations, primarily affecting the central and western parts of the U.S. We have the official confirmed data on when, where, and how many layers were culled from the USDA.[13] By identifying the MSAs where these layers were culled, we pinpoint areas disproportionately affected by the bird flu that might have higher inflation in egg prices early in the outbreak during the inflationary period.

Leveraging this information, we can construct a diff-in-diff identification strategy by grouping treated and control MSAs and comparing the effect of the bird flu outbreak on them. Furthermore, we can use a triple diff-in-diff estimator by additionally interacting MSA-level market concentration with the standard diff-in-diff term to see how the effect varies by 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:

$$P_{mt} = \beta_0 + \beta_1(Treated_m \times Post_t) + \delta_m + \delta_t + \varepsilon_{st}, \tag{5.3}$$

where P_{mt} is the (geometric) Laspeyres inflation rate for eggs in MSA m in quarter t. $Treated_m$ is an indicator variable for whether farmers in MSA m had to cull their layers during the 2014-2015 bird flu 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.

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

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

The results are shown in Table 11. 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 is masking heterogeneity in effects during this period. If we separate the sample into inflationary and deflationary periods, we see opposing effects in the MSAs where layers were culled. In column 2, we restrict our sample to the inflation period when the national eggs inflation rate was above zero, and we estimate a 0.04 coefficient on the interaction of Bird Flu and Post. This 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 teh sample to the deflationary period and see that MSAs that culled their layers experienced a 4 percentage point lower inflation rate after 2014q4. This point estimate is 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 11. The left panel plots the standard event study difference-in-differences coefficients . The dashed vertical line corresponds to 2014q4, which is the start of the post-period. We see no systematic difference in inflation rates between MSAs that culled their layers (the treated MSAs) and MSAs that did not cull their layers (the controlled MSAs) prior to 2014q4.[14] However, after 2014q4, the treated MSAs experienced relatively higher inflation in some quarters and relatively less inflation in other quarters.

This opposing inflation effects can be explained by heterogeneous effects depending on whether there is an inflationary or deflationary period in the egg market. The dependence on inflationary or deflationary period 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. We continue to find no significant difference between these two groups of MSAs prior to 2014q4.

[14] This satisfies the parallel trends assumption.

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

$$
\begin{aligned}
P_{mt} = \beta_0 + \beta_1 HHI_{mt} &+ \beta_2(\text{Treated}_m \times \text{Post}_t) \\
&+ \beta_3(\text{Treated}_m \times HHI_{mt}) + \beta_4(\text{Post}_t \times HHI_{mt}) \\
&+ \beta_5(\text{Treated}_m \times \text{Post}_t \times HHI_{mt}) + \delta_m + \delta_t + \varepsilon_{mt},
\end{aligned}
\tag{5.4}
$$

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 andδ_t, are the same as before, and ε_{st} is the error term.

The results are presented in Table 11. In column 1, we restrict our sample to the inflationary period and find that the MSAs with higher market concentration increased prices at faster rates in the eggs 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 may potentially lower prices by larger amounts during the deflationary period. We find no support for this supposition when we restrict our sample to the deflationary period in column 2. We find some suggestive evidence that these MSAs are slower to decrease prices in the deflationary period, as evidenced by the positive coefficient on the triple interaction term in column 2. This coefficient is significant at the 10% level. In column 3, we pool all quarters in the two-year window together, and we continue to find that MSAs with higher market concentration exhibit higher inflation than those with lower market concentration.

Our results indicate that the market concentration of retailers is a potential mechanism explaining the heterogeneous inflation rates between poor and rich MSAs in the eggs market. In particular, the triple difference-in-differences results suggest that these differences may occur through markups rather than marginal cost with high cost pass-through.[15] Subsequent studies expanding on these analyses will test this hypothesis and disentangle where the effect comes from.

6 Markup Estimation

In this section, we outline our future plan on estimating markups at the retailer level. We aim to follow Hottman (2017) and build a simple model to estimate markups at the MSA level.

6.1 Theoretical Framework

6.1.1 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:

$$U_{mt} = \left[\sum_{s \in S_{mt}} (\varphi_{smt} C_{smt})^{\frac{\sigma_S - 1}{\sigma_S}} \right]^{\frac{\sigma_S}{\sigma_S - 1}}, \tag{6.5}$$

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

[15] If differences in marginal costs are the main driving factor behind the heterogeneous inflation rates, we would expect to see higher deflation in the deflationary period in MSAs with higher market concentration.

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:

$$C_{smt} = \left[\sum_{i \in I_{smt}} \left(\varphi_{ismt} C_{ismt} \right)^{\frac{\sigma_I - 1}{\sigma_I}} \right]^{\frac{\sigma_I}{\sigma_I - 1}}, \tag{6.6}$$

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:

$$C_{ismt} = \left[\sum_{u \in U_{ismt}} \left(\varphi_{usmt} C_{usmt} \right)^{\frac{\sigma_U - 1}{\sigma_U}} \right]^{\frac{\sigma_U}{\sigma_U - 1}}, \tag{6.7}$$

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:

$$\left(\Pi_{u \in U_{ismt}} \varphi_{usmt} \right)^{\frac{1}{N_{ismt}}} = 1 \tag{6.8}$$

$$\left(\Pi_{i \in I_{smt}} \varphi_{ismt} \right)^{\frac{1}{N_{smt}}} = 1, \tag{6.9}$$

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.

6.1.2 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:

$$S_{usmt} = \frac{(P_{usmt}/\varphi_{usmt})^{1-\sigma_U}}{\sum_{k\in U_{ismt}}(P_{kist}/\varphi_{kist})^{1-\sigma_U}}, \tag{6.10}$$

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:

$$P_{ismt} = \left[\sum_{k\in U_{ismt}} \left(\frac{P_{ksmt}}{\rho_{ksmt}}\right)^{1-\sigma_U}\right]^{\frac{1}{1-\sigma_U}}. \tag{6.11}$$

6.1.3 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:

$$S_{ismt} = \frac{(P_{ismt}/\varphi_{ismt})^{1-\sigma_I}}{\sum_{k\in I_{smt}}(P_{ismt}/\varphi_{ismt})^{1-\sigma_I}}, \tag{6.12}$$

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:

$$P_{smt} = \left[\sum_{k\in G_{smt}} \left(\frac{P_{ksmt}}{\rho_{ksmt}}\right)^{1-\sigma_I}\right]^{\frac{1}{1-\sigma_I}}. \tag{6.13}$$

6.1.4 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:

$$S_{smt} = \frac{(P_{smt}/\varphi_{smt})^{1-\sigma_S}}{\sum_{k \in S_{mt}} (P_{kt}/\varphi_{kt})^{1-\sigma_S}}, \tag{6.14}$$

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:

$$P_{mt} = \left[\sum_{k \in S_{mt}} \left(\frac{P_{kmt}}{\rho_{kmt}} \right)^{1-\sigma_S} \right]^{\frac{1}{1-\sigma_S}}. \tag{6.15}$$

6.1.5 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:

$$E_{usmt} = S_{usmt} S_{ismt} S_{smt} E_{mt}. \tag{6.16}$$

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

$$Q_{usmt} = \frac{E_{usmt}}{P_{usmt}} \tag{6.17}$$

Rephrasing it with (6.10), (6.12), (6.14), and (6.16), we have the following:

$$Q_{usmt} = \varphi_{usmt}^{\sigma_U - 1} \varphi_{ismt}^{\sigma_I - 1} \varphi_{smt}^{\sigma_s - 1} P_{usmt}^{-\sigma_U} P_{ismt}^{\sigma_U - \sigma_I} P_{smt}^{\sigma_I - \sigma_s} P_{smt}^{\sigma_s - 1} E_{mt}. \tag{6.18}$$

6.1.6 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:

$$V_{usmt}(Q_{usmt}) = z_{usmt} Q_{usmt}^{1+\delta_i}, \tag{6.19}$$

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:

$$\pi_{rmt} = \sum_{u \in U_{rmt}} P_{urmt} Q_{urmt} - V_{urmt}(Q_{urmt}) - F_{mt}, \tag{6.20}$$

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:

$$Q_{usmt} + \sum_{k \in U_{rmt}} \left(P_{ksmt} \frac{\partial Q_{ksmt}}{\partial P_{ksmt}} - \frac{\partial V_{ksmt}(Q_{ksmt})}{\partial Q_{ksmt}} \frac{\partial Q_{ksmt}}{\partial P_{ksmt}} \right) = 0. \tag{6.21}$$

The optimal price is then given by

$$P_{usmt} = \mu_{rmt} m_{usmt}, \tag{6.22}$$

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

$$\mu_{rmt} = \frac{\epsilon_{rmt}}{\epsilon_{rmt}-1}, \tag{6.23}$$

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

$$\epsilon_{rmt} = \sigma_s - (\sigma_s - 1)S_{rmt}, \tag{6.24}$$

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.[16]

6.2 Structural Estimation

Now, we outline how to estimate the structural model given the demand and the elasticity of substitution at each tier. Once the elasticity of substitution is obtained, we can back out the demand elasticity in (6.24) and markup in (6.23) eventually.

6.2.1 Lowest Tier: Elasticity of Substitution across UPCs

In this step, estimating σ_U follows the approach in Feenstra (1994), Broda and Weinstein (2010), and Hottman (2017). The identification strategy is as follows. For a given food item i, the slope of the demand and supply, σ_U and δ_i, are assumed to be constant across barcodes and time, but their intercepts are allowed to vary across barcodes and time.

[16] 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 $\frac{\sigma_s}{\sigma_s - 1}$.

Starting with (6.10), we take the time difference and the difference of the result relative to other barcode within the same food item and store. The double difference term is given by:

$$\Delta^{k,t} \ln S_{usmt} = (1 - \sigma_U)\Delta^{k,t} \ln P_{usmt} + \varepsilon_{usmt}, \tag{6.25}$$

where $\Delta^{k,t} x$ implies the double difference of x between time t and $t-1$ and with respect to other UPC k; $\varepsilon_{usmt} = (1 - \sigma_U)(\Delta^t \ln \varphi_{ksmt} - \Delta^t \ln \varphi_{usmt})$ is the error term.

Next, we use and double difference the pricing equation (6.22), which gives:

$$\Delta^{k,t} \ln P_{usmt} = \frac{\delta_i}{1 + \delta_i}\Delta^{k,t} \ln S_{usmt} + \omega_{usmt}, \tag{6.26}$$

where the unobserved error term is $\omega_{usmt} = \frac{1}{1+\delta_i}(\Delta^t \ln z_{usmt} - \Delta^t \ln z_{ksmt})$.

The orthogonality condition for each barcode is then defined as

$$G(\beta_i) = \mathbb{E}_T[\chi_{usmt}(\beta_i)] = 0, \tag{6.27}$$

where $\beta_i = \begin{bmatrix} \sigma_U \\ \delta_i \end{bmatrix}$ and $\chi_{usmt} = \varepsilon_{usmt}\omega_{usmt}$.[17]

Note that this condition assumes the orthogonality of the idiosyncratic demand and supply shocks at the barcode level, after barcode and quarter fixed effects have been differenced out. This orthogonality is plausible because product characteristics are fixed for each barcode and advertising typically occurs at the level of the brand. Supply shocks such as labor strikes or changes in manufacturing costs are unlikely to be correlated with quarterly demand shocks at the store level.

For each food item, we form the GMM objective function objective function as follows:

$$\hat{\beta}_i = \underset{\beta_i}{\operatorname{argmin}} \left\{ G^*(\beta_i)' W G^*(\beta_i) \right\}, \tag{6.28}$$

where $G^*(\beta_i)$ is the stacked set of $G(\beta_i)$ over all barcodes in food i and W is positive definite weighting matrix. Following Broda and Weinstein (2010), I give more weights

[17] See further details in Appendix D.1.

to barcodes that are present for longer time periods in the data. With the estimates of σ_U, we construct food item price indices in (6.11).

6.2.2 Middle Tier: Elasticity of Substitution across Food Items

Next, we time difference (6.12) and difference it relative to another food item in the same store s to obtain

$$\Delta^{i,t} \ln S_{ismt} = (1 - \sigma_I)\Delta^{i,t} \ln P_{ismt} + \varepsilon_{ismt}, \tag{6.29}$$

where the unobserved error term is $\varepsilon_{ismt} = -(\sigma_I - 1)\Delta^{i,t} \ln \varphi_{ismt}$.

Note that the ordinary least squares (OLS) estimation of (6.29) is expected to be biased as the error term is likely correlated with the double-differenced food item price index. For instance, if there is a relative increase in food item quality, this can raise the quantity demeaned of the barcodes within the food item and thus raise the food item price index with upward-sloping barcode supply. Thus, we need an instrumental variable to fix it and follow the approach in Hottman et al. (2016).

Based on the normalization of quality as in (6.8), we can obtain the double-differenced CES food item price index as follows:

$$\Delta^{i,t} \ln P_{ismt} = \Delta^{i,t} \ln \tilde{P}_{ismt} + \frac{1}{1 - \sigma_U} \Delta^{i,t} \ln \left(\sum_{u \in U_{ismt}} \frac{S_{usmt}}{\tilde{S}_{ismt}} \right), \tag{6.30}$$

where $\tilde{X}_{ismt} \equiv \left(\Pi_{k \in U_{ismt}} X_{ksmt} \right)^{\frac{1}{N_{ismt}}}$ indicates the geometric mean of X across barcodes within food item i and store s at time t. See the derivation in Appendix D.2.

We estimate σ_I by using the second term in (6.30) as an instrument for the food item price index in (6.29).[18] The moment condition for instrumental variables is

$$\mathbb{E}\left[\varepsilon_{ismt}\Delta^{i,t}\ln\left(\sum_{u\in U_{ismt}}\frac{S_{usmt}}{\tilde{S}_{ismt}}\right)\right]=0. \tag{6.31}$$

As before, with an estimate of σ_I, we construct store price indices following (6.13).

6.2.3 Upper Tier: Elasticity of Substitution across Stores

Lastly, we time difference (6.14) and difference it relative to another store within the same chain and MSA. This gives us the following term:

$$\Delta^{s,t}\ln S_{smt}=(1-\sigma_S)\Delta^{s,t}\ln P_{smt}+\varepsilon_{smt}, \tag{6.32}$$

where the error term is $\varepsilon_{smt}=-(\sigma_S-1)\Delta^{s,t}\ln\varphi_{smt}$.

As in the previous subsection, estimating σ_S requires an instrumental variable. Based on the normalization in (6.9), we can double-difference the store price index (6.15) as follows:

$$\begin{aligned}\Delta^{s,t}\ln P_{smt} &= \frac{1}{1-\sigma_I}\Delta^{s,t}\ln\left(\sum_{i\in I_{smt}}\frac{S_{ismt}}{\tilde{S}_{smt}}\right)+\Delta^{s,t}\left\{\frac{1}{N_{smt}}\sum_{i\in I_{smt}}\left(\frac{1}{1-\sigma_U}\ln\left(\sum_{u\in U_{ismt}}\frac{S_{usmt}}{\tilde{S}_{ismt}}\right)\right)\right\}\\ &+ \Delta^{s,t}\left\{\frac{1}{N_{smt}}\sum_{i\in I_{smt}}\ln\tilde{P}_{ismt}\right\}.\end{aligned} \tag{6.33}$$

See the derivation in Appendix D.3.

We estimate σ_S using the sum of the first two terms in (6.33) as an instrument for the store price index in (6.32). The moment condition for instrument variables is

$$\mathbb{E}\left[\varepsilon_{smt}\Delta^{s,t}\left\{\ln\left[\sum_{i\in I_{smt}}\frac{S_{ismt}}{\tilde{S}_{smt}}\right]+\frac{1}{N_{smt}}\sum_{i\in I_{smt}}\frac{1}{1-\sigma_U}\ln\left[\sum_{u\in U_{ismt}}\frac{S_{usmt}}{\tilde{S}_{ismt}}\right]\right\}\right]=0. \tag{6.34}$$

[18] See further discussion in Hottman (2021).

7 Concluding Remarks

We document that poorer MSAs in the US experienced higher inflation rates than the richest MSAs in the US for both aggregated food and disaggregated food categories between 2006 and 2020. This finding is also robust to different types of price indexes and to the set of common goods consumed across all MSA deciles. Furthermore, we document that official price indexes PCE systematically understate the inflation that poorer areas experience by having price indexes closer to the richest decile.

We investigate how this pattern is linked to different retailer dynamics across poor and rich MSAs, finding that the composition and market concentration of retailers vary across different regions. In particular, we find a positive association between retailer sales concentration and inflation rates. To develop a more causal link between market concentration and inflation, we exploit the 2014-2015 bird flu episode and employ a triple-difference estimator. We examine the MSAs affected by the bird flu and find that those with higher HHI values experience higher inflation rates than those with lower HHI values. This result suggests that retailer market concentration plays a role when retailers face a cost shock by charging higher prices.

This work is the first step in a larger research agenda. We would like to further investigate whether the inflation gap across MSAs is linked to retailers' market power by estimating markups in the data as outlined in Section 6. Additionally, we would like to build a structural model to quantify the channel and derive more testable and policy-related implications. Lastly, we will continue to analyze the unrepresentativeness of official price indexes.

References

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.

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.

Feenstra, Robert C, Luca Macedoni, and Mingzhi Xu, "Large Firms, Consumer Heterogeneity and the Rising Share of Profits," Working Paper 29646, National Bureau of Economic Research January 2022.

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.

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

_ , Stephen J. Redding, and David E. Weinstein, "Quantifying the Sources of Firm Heterogeneity," *Quarterly Journal of Economics*, 03 2016, *131* (3), 1291–1364.

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 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).

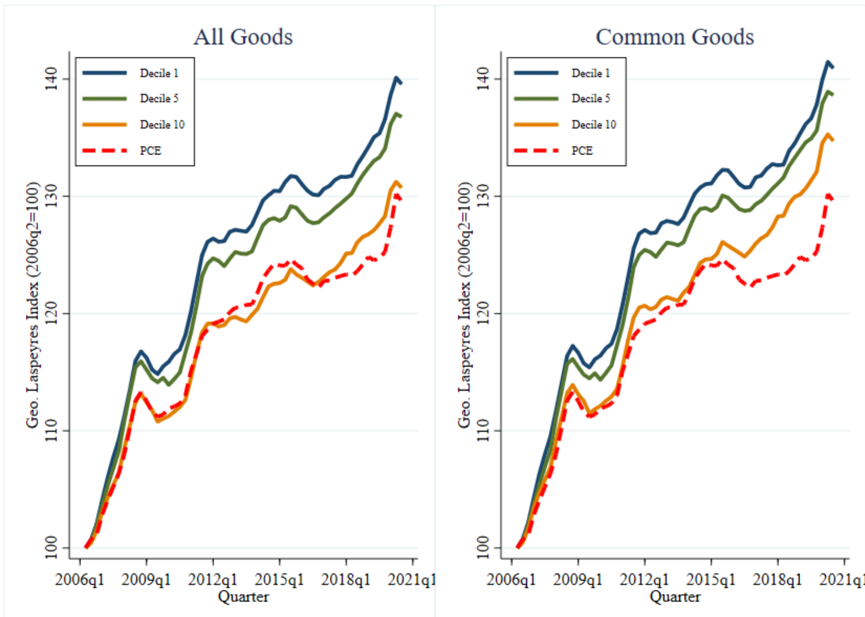
Moretti, Enrico, "Real Wage Inequality," *American Economic Journal: Applied Economics*, January 2013, *5* (1), 65–103.

Nevo, Aviv, "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

Weyl, E. Glen and Michal Fabinger, "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy*, 2013, *121* (3), 528–583.

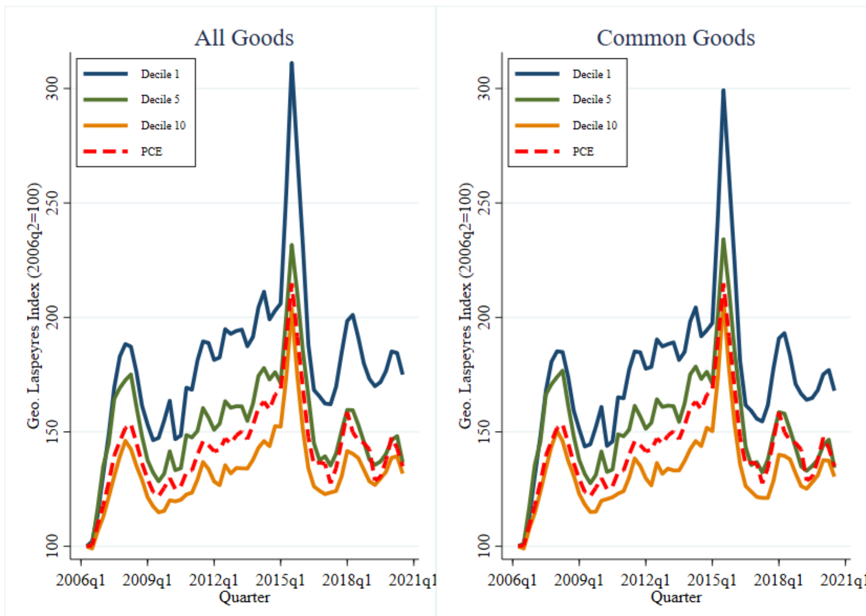
A Figures

Figure 1: Price Index for Aggregated Food



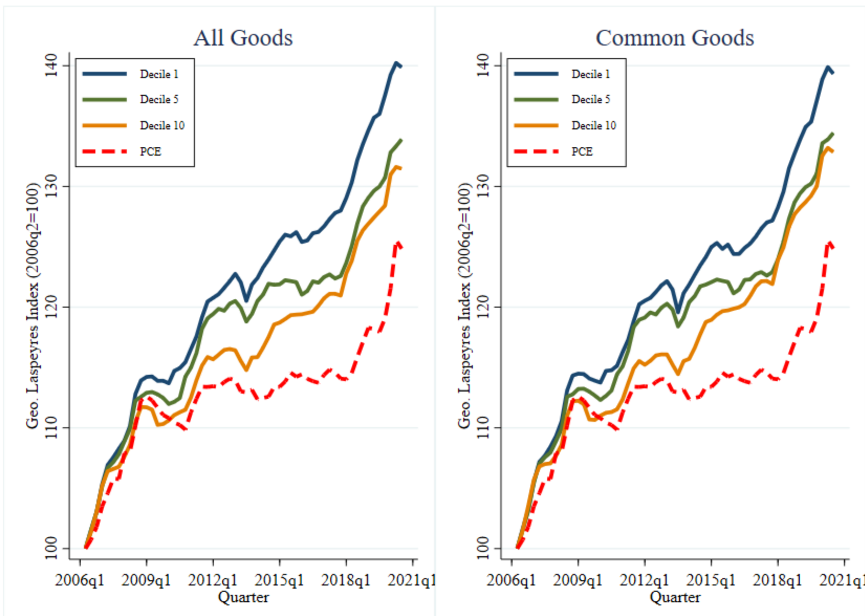
Note: This figure represents relative prices for the aggregated food market with four series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2020Q4. The data for the three solid lines come from the NielsenIQ Retail Scanner dataset, represented 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 2: Laspeyres Price Index for Eggs



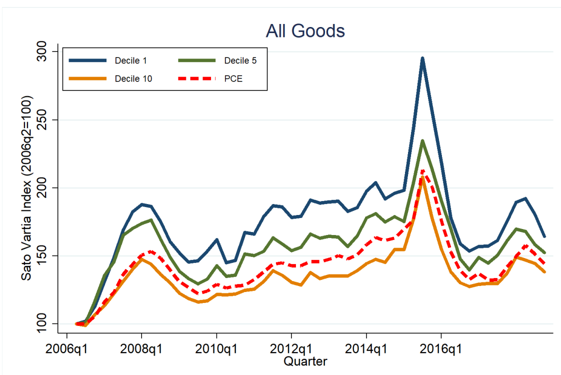
Note: The figure represents relative prices in the aggregated egg market with five series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2020Q4. The data come from NielsenIQ Retail Scanner dataset represented by geometric Laspeyres price 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 3: Laspeyres Price Index for Soda and Juices

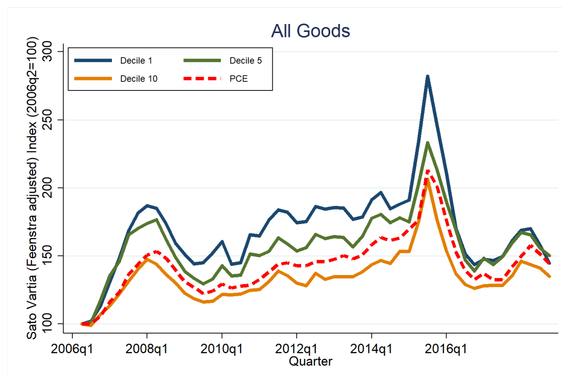


Note: The figure represents relative prices for the soda and juices market with five series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2020Q4. The data come from NielsenIQ Retail Scanner dataset represented by geometric Laspeyres 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.

Figure 4: Demand-based Price Indexes for Eggs



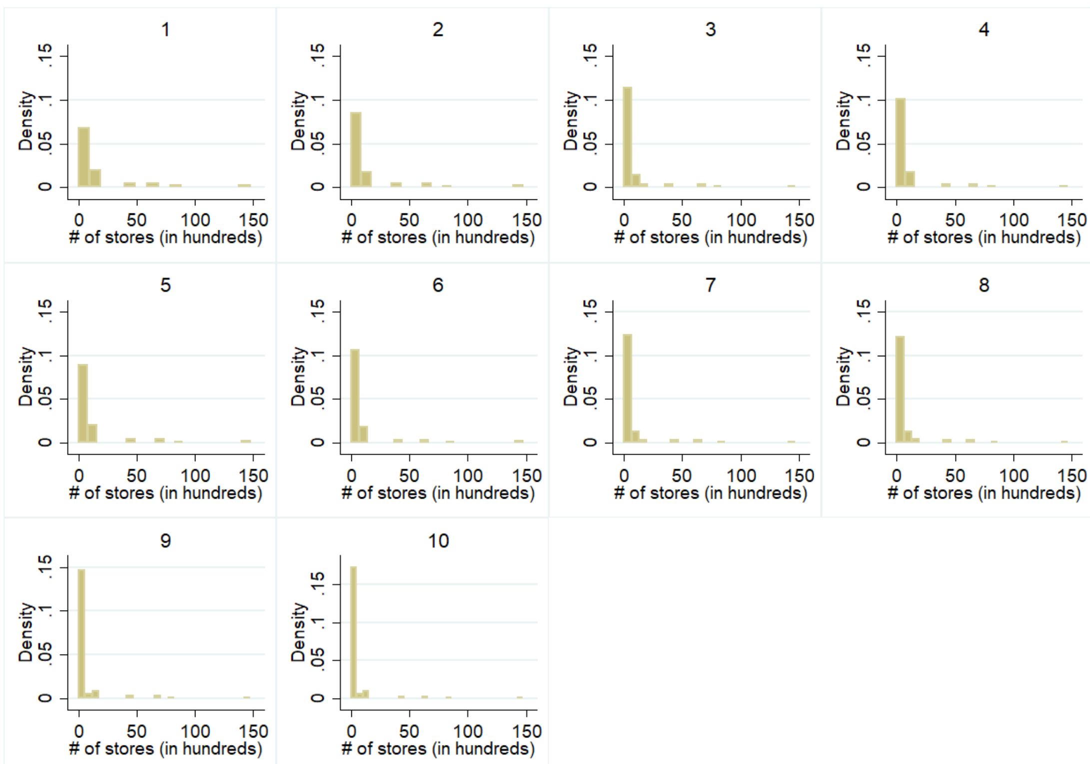
Sato Vartia



Feenstra-adjusted Sato Vartia

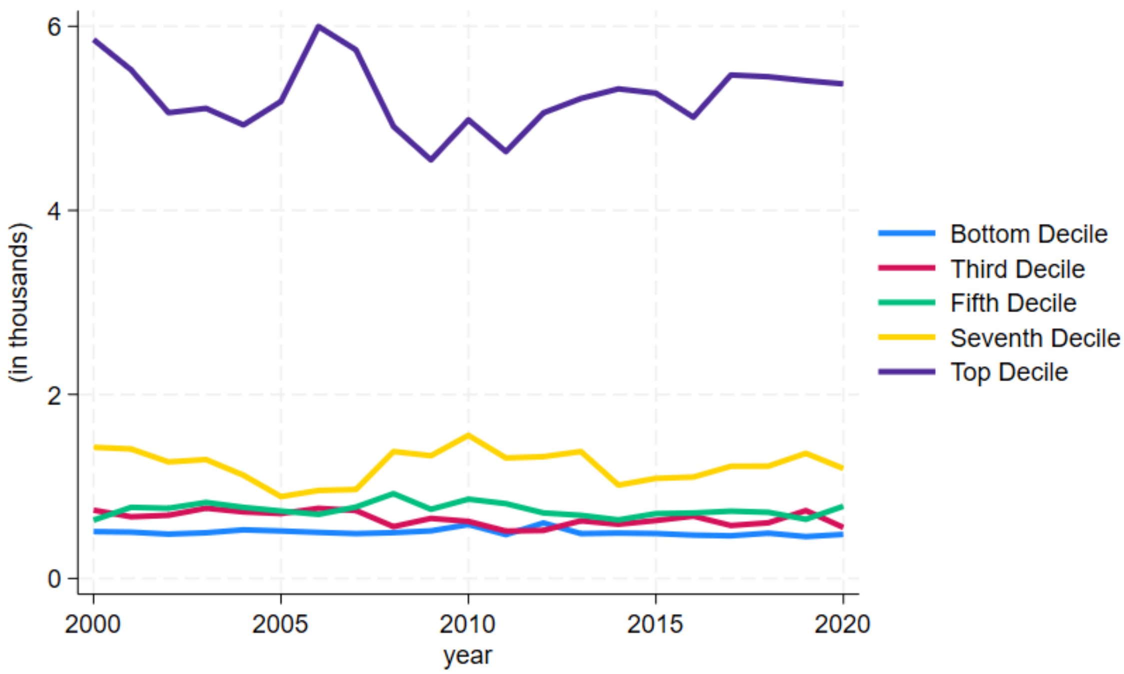
Note: The figure represents relative prices for the aggregated egg market with five series, where each series is normalized to 100 at the start of the sample. The sample period begins in 2006Q2 and ends in 2016Q4. The data come from NielsenIQ Retail Scanner dataset represented by Sato-Vartia and Feenstra-adjusted 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.

Figure 5: Distribution of Retailer Size by Income Decile



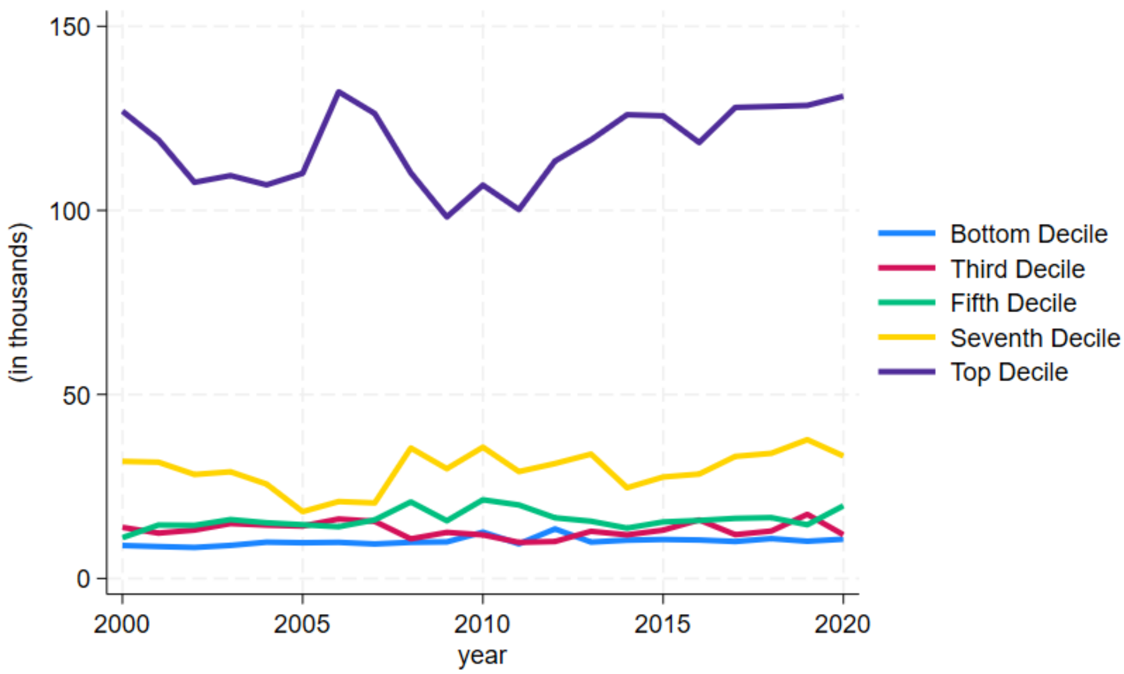
Note: The figure shows the distribution of the number of stores (at the national level) of chains located in each income decile. The data come from NielsenIQ for the aggregate food and beverages.

Figure 6: Number of Retail Chains by Income Decile



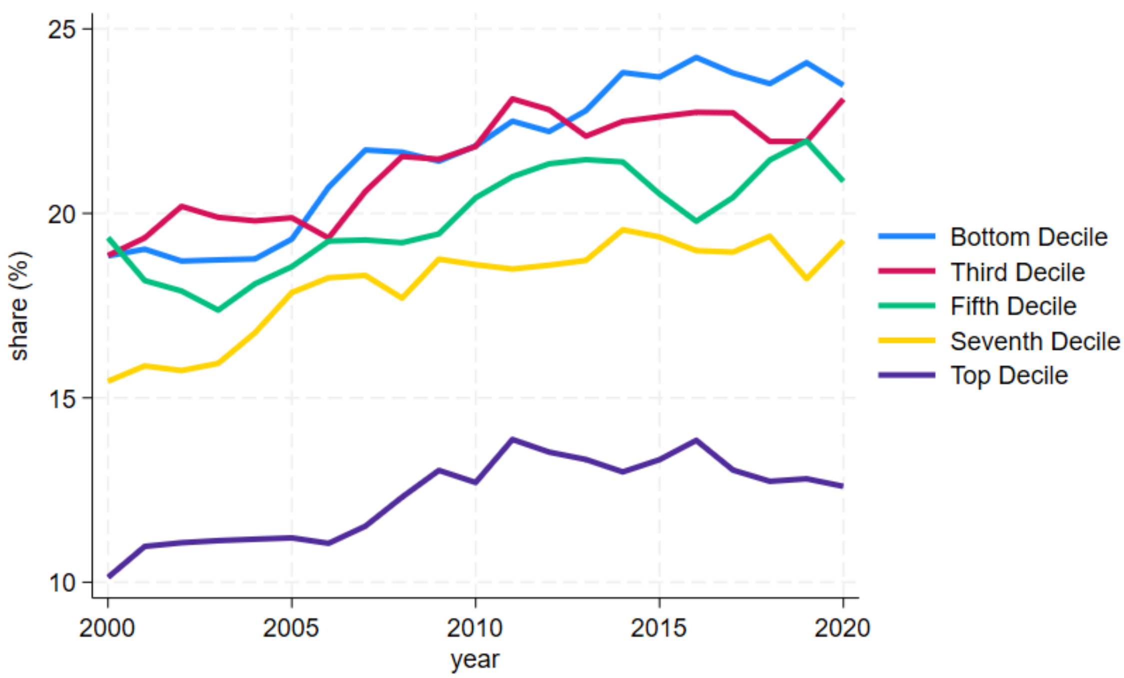
Note: The figure represents the number of retail chains present from 2000 to 2020 in deciles 1, 3, 5, 7, and 10. The data on the number of retail chains come from the Business Dynamics Statistics (BDS). We specifically only use data on chains from the BDS for the retail trade sector (NAICS codes 44-45).

Figure 7: Employment of Retail Chains by Income Decile



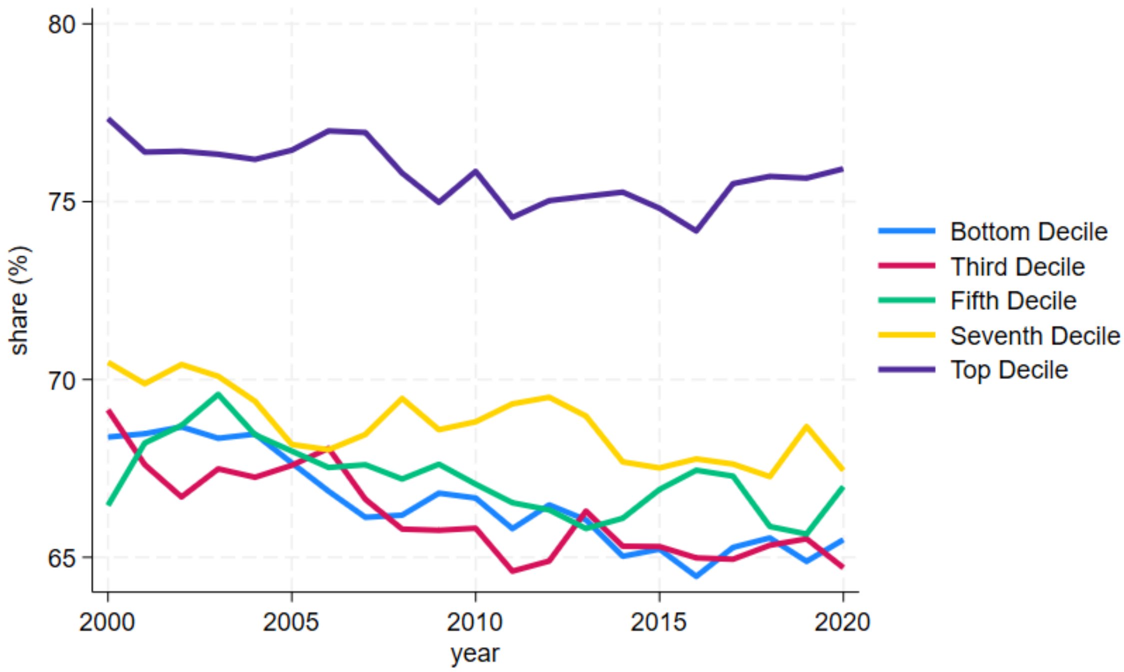
Note: The figure represents the employment of retail chains present from 2000 to 2020 in deciles 1, 3, 5, 7, and 10. The data on the number of retail chains come from the Business Dynamics Statistics (BDS). We specifically only use data on chains from the BDS for the retail trade sector (NAICS codes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

Figure 8: Share of Large Retailers



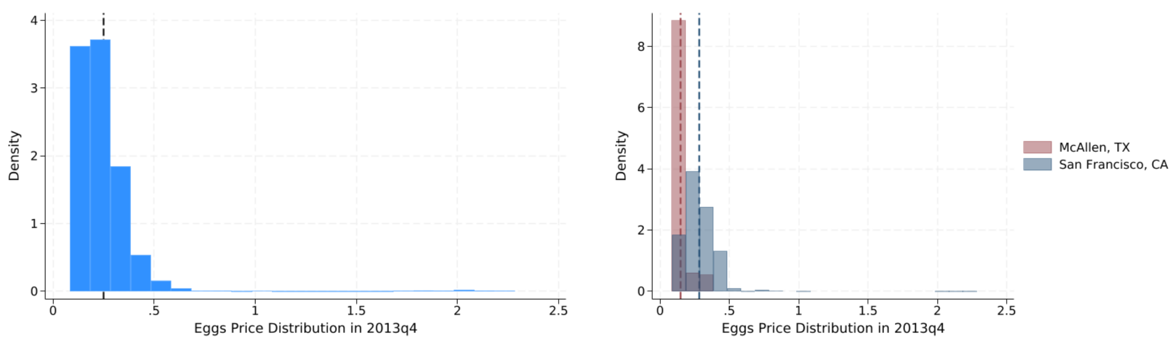
Note: The figure represents the share of establishments from large retail chains present from 2006 to 2019 in deciles 1, 3, 5, 7, and 10. A large retail chain is defined as having (at the firm level) more than 500 employees. The data on the number of retail chains come from the Business Dynamics Statistics (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 9: Share of Small Retailers



Note: 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).

Figure 10: Price Levels of Eggs (UPCs)

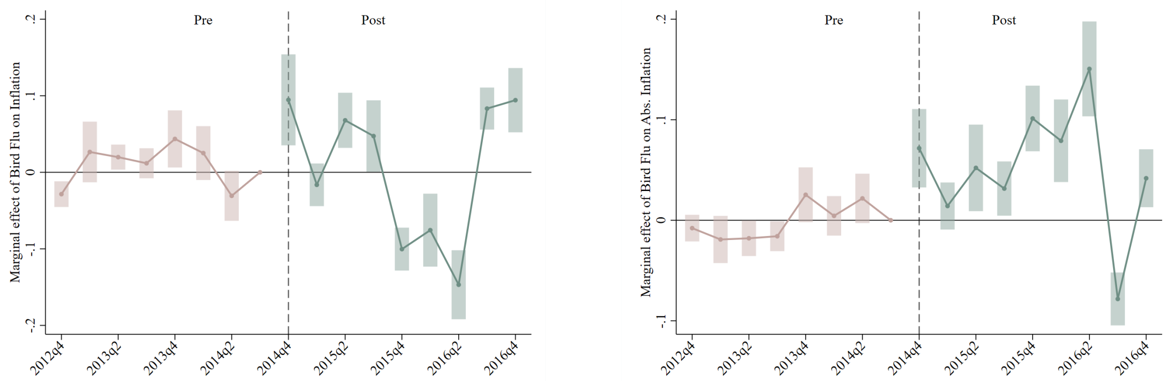


(a) Entire Sample

(b) Two MSAs

Note: The figure represents the distribution of prices of UPCs of the PCE category "eggs" in 2013q4. Prices are denominated in per-egg terms. The figure on the left corresponds to all areas covered by NielsenIQ, while the figure on the right corresponds to the poorest (McAllen, TX) and richest MSAs in our sample (San Francisco, CA). The vertical dashed line indicates the median price.

Figure 11: Event Study Difference-in-Differences (Bird Flu)



(a) Treated × Quarter

(b) Treated × Quarter (Absolute Value)

Note: 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.

B Tables

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		

Note: 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.

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.

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

	Decile 1	Decile 3	Decile 5	Decile 7	Decile 10
	Mean	Mean	Mean	Mean	Mean
	(SD)	(SD)	(SD)	(SD)	(SD)
Income per capita ($1,000)	31.37	36.24	38.93	42.18	57.14
	(5.28)	(4.40)	(4.92)	(5.19)	(11.48)
Sales ($1mil.)	26.427	60.86	79.54	136.22	730.63
	(25.37)	(59.57)	(113.72)	(203.59)	(745.44)
Number of chains	7.76	8.72	8.80	9.42	13.16
	(3.06)	(2.76)	(2.96)	(3.37)	(4.96)
Number of stores	61.76	84.91	99.74	164.79	511
	(50.44)	(63.25)	(97.72)	(171.85)	(479.69)
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)
Market concentration	0.37	0.36	0.35	0.32	0.20
	(0.16)	(0.16)	(0.14)	(0.13)	(0.11)

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 measured by the HHI (Herfindahl-Hirschman Index) of chain-level sales.

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

*** p<0.01, ** p<0.05, * p<0.1

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.

Table 6: 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

*** p<0.01, ** p<0.05, * p<0.1

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 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.

Table 7: HHI across Different Income Deciles

	HHI
Decile	-0.004***
	(0.000)
Year FE	Yes
Item FE	Yes
Observations	10,920
*** p<0.01, ** p<0.05, * p<0.1	

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).

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".

Table 9: Price Dispersion in Eggs Market

	Price	Price
Income per capita ($10k)	0.01***	0.006***
	(0.000)	(0.000)
UPC FE	Yes	Yes
Quarter FE	Yes	Yes
Retailer FE	No	Yes
Observations	11,232,109	11,231,937

Note: This table shows regression results where the unit of observation is the UPC-store-quarter level. The data used are UPCs from the PCE category "eggs" from 2006q1 to 2020q4. The price of eggs, on a per-egg basis, is the dependent variable. The independent variable is the average income per capita of an MSA, which is denominated in $10,000. The first column imposes UPC and quarter fixed effects. The second column adds retailer fixed effects. Standard errors are clustered at the MSA level.

Table 10: Market Concentration in Eggs Market

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

*** p<0.01, ** p<0.05, * p<0.1

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.

Table 11: TWFE Estimator (Bird Flu Episode)

	Inflation	Inflation	Inflation	Abs. Inflation
Bird Flu $\times$ 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

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

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.

Table 12: 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

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

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.

C Robustness with the Consumer Panel

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 13 show that, on average, 92% of households made purchases exclusively within their residential MSAs.

Table 13: 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

Note: 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.

Furthermore, when examining household characteristics and shopping patterns by each category, Table 14 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 15 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

Table 14: 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

Note: 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.

our baseline measures of income deciles based on store locations and BEA income per capita data are not mismeasured.

Table 15: 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

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

D Model Derivation

In this section, we provide further details behind the estimation strategy and process in Section 6.2.

D.1 Derivation of (6.25) and (6.26) in the Lowest Tier

First, using (6.10) and taking a log, we have

$$\ln S_{usmt} = (1-\sigma_U)\ln P_{usmt} - (1-\sigma_U)\ln \varphi_{ismt} + \ln \sum_{k\in I_{ismt}} \left(\frac{P_{ksmt}}{\varphi_{ksmt}}\right)^{1-\sigma_U}. \tag{D.35}$$

Note that the last term is constant across barcodes. Therefore, double-differencing (D.35) with respect to time and a baseline barcode gives (6.25).

Furthermore, using (6.22) and $S_{usmt} = P_{usmt}Q_{usmt}$ and taking a log, we have

$$\ln P_{usmt} = \frac{1}{1+\delta_i}\Bigg(\ln \mu_{rmt} + \ln(1+\delta_i)\Bigg) + \frac{1}{1+\delta_i}\ln z_{usmt} + \frac{\delta_i}{1+\delta_i}\ln S_{usmt}. \tag{D.36}$$

Note that the first term is constant across barcodes and time, so the double-differencing gives (6.26) from the remaining terms.

Lastly, multiplying these equations gives us (6.27), which can be rephrased as the following regression model:

$$\Bigg(\Delta^{k,t}\ln P_{usmt}\Bigg)^2 = \theta_1\Bigg(\Delta^{k,t}\ln S_{usmt}\Bigg)^2 + \theta_2\Bigg(\Delta^{k,t}\ln P_{usmt}\Bigg)\cdot\Bigg(\Delta^{k,t}\ln S_{usmt}\Bigg) + \chi_{usmt}, \tag{D.37}$$

where

$$\theta_1 = \frac{\delta_i}{(1+\delta_i)(\sigma_U-1)} \tag{D.38}$$

$$\theta_2 = \frac{1-\delta_i(\sigma_U-2)}{(1+\delta_i)(\sigma_U-1)} \tag{D.39}$$

$$\chi_{usmt} = \varepsilon_{usmt}\omega_{usmt}. \tag{D.40}$$

D.2 Derivation of (6.29) and (6.30) in the Middle Tier

First, using (6.12) and taking a log, we have

$$\ln S_{ismt} = (1-\sigma_I)\ln P_{ismt} - (1-\sigma_I)\ln\varphi_{ismt} + \ln\sum_{k\in I_{smt}}\left(\frac{P_{ksmt}}{\varphi_{ksmt}}\right)^{1-\sigma_I}, \tag{D.41}$$

where the last term is constant across food item i. Therefore, double-differencing (D.41) with respect to time and a baseline item gives (6.29).

With an estimate of σ_U, we revisit (6.11) and construct the price index for each food item i sold in store s (in MSA m) at time t. Given (6.10), we can rephrase it as follows:

$$\begin{aligned} P_{ismt} &= \left[\sum_{k\in U_{ismt}}\left(S_{ksmt}\sum_{k\in U_{ismt}}(P_{kist}/\varphi_{kist})^{1-\sigma_U}\right)\right]^{\frac{1}{1-\sigma_U}} \\ &= \left[\left(\sum_{k\in U_{ismt}}S_{ksmt}\right)\left(\sum_{k\in U_{ismt}}(P_{kist}/\varphi_{kist})^{1-\sigma_U}\right)\right]^{\frac{1}{1-\sigma_U}} \end{aligned}$$

Taking a log, we obtain

$$\ln P_{ismt} = \frac{1}{1-\sigma_U}\left[\ln\left(\sum_{k\in U_{ismt}}S_{ksmt}\right) + \ln\left(\sum_{k\in U_{ismt}}(P_{kist}/\varphi_{kist})^{1-\sigma_U}\right)\right] \tag{D.42}$$

As another step, using (6.10) and the normalization imposed in (6.8), we take the geometric mean of the UPC-level price across barcodes within each food item i and store s at time t get the following term:

$$\begin{aligned} \ln\tilde{P}_{ismt} &= \ln\left(\Pi_{k\in U_{ismt}}P_{ksmt}^{\frac{1}{N_{ismt}}}\right) = \frac{1}{N_{ismt}}\sum_{k\in U_{ismt}}\ln P_{ksmt} \\ &= \frac{1}{1-\sigma_U}\left[\frac{1}{N_{ismt}}\sum_{k\in U_{ismt}}\ln S_{usmt} + \ln\left(\sum_{k\in U_{ismt}}(P_{kist}/\varphi_{kist})^{1-\sigma_U}\right)\right] \end{aligned}$$

$$= \frac{1}{1-\sigma_U}\left[\ln \tilde{S}_{ismt} + \ln\left(\sum_{k\in U_{ismt}} (P_{kist}/\varphi_{kist})^{1-\sigma_U}\right)\right]. \tag{D.43}$$

Combining (D.42) and (D.43), we have

$$\ln\left(\frac{P_{ismt}}{\tilde{P}_{ismt}}\right) = \frac{1}{1-\sigma_U}\ln\left(\frac{\sum_{k\in U_{ismt}} S_{ksmt}}{\tilde{S}_{ismt}}\right),$$

and thus,

$$\ln P_{ismt} = \ln \tilde{P}_{ismt} + \frac{1}{1-\sigma_U}\ln\left(\frac{\sum_{k\in U_{ismt}} S_{ksmt}}{\tilde{S}_{ismt}}\right). \tag{D.44}$$

Finally, double-differencing (D.44) gives (6.30).

D.3 Derivation of (6.32) and (6.33) in the Top Tier

Similarly as before, with an estimate of σ_I and (6.12), we can rephrase (6.13) as follows:

$$\ln P_{smt} = \frac{1}{1-\sigma_I}\left[\ln\left(\sum_{k\in I_{smt}} S_{ksmt}\right) + \ln\left(\sum_{k\in I_{smt}} (P_{ksmt}/\varphi_{ksmt})^{1-\sigma_I}\right)\right] \tag{D.45}$$

As another step, using (6.12) and the normalization imposed in (6.9), we take the geometric mean of the item-level price across food items within each store s at time t get the following term:

$$\ln \tilde{P}_{smt} = \frac{1}{1-\sigma_I}\left[\ln \tilde{S}_{smt} + \ln\left(\sum_{k\in I_{smt}} (P_{ksmt}/\varphi_{ksmt})^{1-\sigma_I}\right)\right]. \tag{D.46}$$

Combining (D.45) and (D.46), we have

$$\ln\left(\frac{P_{smt}}{\tilde{P}_{smt}}\right) = \frac{1}{1-\sigma_I}\ln\left(\frac{\sum_{k\in I_{ksmt}} S_{smt}}{\tilde{S}_{smt}}\right),$$

and thus,

$$\ln P_{smt} = \ln \tilde{P}_{smt} + \frac{1}{1-\sigma_I} \ln\left(\frac{\sum_{k\in I_{smt}} S_{ksmt}}{\tilde{S}_{smt}}\right). \tag{D.47}$$

Here, using (D.44) and the definition of

$$\ln \tilde{P}_{smt} = \frac{1}{N_{smt}} \sum_{k\in I_{smt}} \ln P_{ismt},$$

we can rephrase (D.47) further as follows:

$$\ln P_{smt} = \frac{1}{N_{smt}} \sum_{k\in I_{smt}} \ln \tilde{P}_{ksmt} + \frac{1}{N_{smt}} \sum_{k\in I_{smt}} \left(\frac{1}{1-\sigma_U} \ln\left(\frac{\sum_{k\in U_{ismt}} S_{ksmt}}{\tilde{S}_{ismt}}\right)\right) + \frac{1}{1-\sigma_I} \ln\left(\frac{\sum_{k\in I_{smt}} S_{ksmt}}{\tilde{S}_{smt}}\right). \tag{D.48}$$

Finally, double-differencing (D.48) gives (6.33).