Procyclical Markups: Nonparametric Estimates in a Structural Dynamic Panel*

Bulat Gafarov, Tengda Gong, Jens Hilscher

March 18, 2025

Abstract

This paper identifies causes of procyclical variation in markups using micro data. We proceed in two steps. First, we show that markups can be nonparametrically estimated based on a locally linear demand model. We apply a revenue-weighted instrumental variable approach to estimate region-specific time-varying markups using U.S. grocery price scanner data from 2001-2022. We find a secular trend rise and procyclical variation in markups. Second, using a dynamic panel model, we quantify the contributions of four main drivers of variation: demographic trends, market concentration, aggregate demand, and real interest rate. Taken together, these factors explain a large fraction of the variation across markets and over time. The variation of demographic and market concentration factors had a small impact and were moving in opposite directions. From 2008-2016 negative real interest rates increased markups counteracting downward pressure from aggregate demand factors; later in the sample markups increase because of higher aggregate demand.

JEL codes: C3, E43, L13, Q11

Keywords: nonparametric demand estimation, markups, market concentration, grocery store prices, dynamic panels, monetary policy, COVID, aggregate demand, historical decomposition.

*Bulat Gafarov and Jens Hilscher are faculty members from the Department of Agricultural and Resource Economics at the University of California, Davis. Tengda Gong is an Assistant Professor from the Department of Agribusiness and Agricultural Economics at the University of Manitoba. Contacts: bgafarov@ucdavis.edu, tengda.gong@umanitoba.ca, jhilscher@ucdavis.edu. We would like to thank Saroj Bhattarai, Steve Cecchetti, Jim Chalfant, Thibault Fally, Leonard Goff, Oscar Jorda, Sanjay Singh, Konstantin Styrin, and Hal Varian as well as participants at the 2022 AAEA meetings in Anaheim, the Agricultural Economics Seminar at UC Davis (October 2023), the Macro brownbag at UC Davis (November 2023), the 2023 Latin American Meeting of the Econometric Society, Bogota, Colombia, and the 2024 ASSA meetings in San Antonio for helpful comments and Hovhannes Mnatsakanyan for research assistance. A previous version of the paper was circulated under the title "Secular Rise and Pro-Cyclical Variation in Markups: Evidence from US Grocery Stores." Research for this paper was supported by a grant from the Giannini Foundation of Agricultural Economics. Part of the scanner data used in this paper comes from NielsenIQ, part from IRI. Disclaimers: (i) Researchers' 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 researchers 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. (ii) All estimates and analyses in this paper, based on data provided by IRI, are by the authors and not by IRI.

1 Introduction

Does policy intervention supporting aggregate demand increase competition and welfare? To answer this question, one can measure changes in price markups over marginal costs—if markups decrease, then competition and welfare increase. We can thus rephrase our question: do markups decrease when aggregate demand increases? Are markups countercyclical? This hypothesis has been present at least since Keynes (1936). In his *General Theory*, he hypothesized that markups increase during economy-wide downturns when aggregate demand declines. The same aggregate pattern is also present in the modern New Keynesian models, e.g. Broer et al. (2020), and models of oligopolistic price wars in the absence of sticky prices (Rotemberg and Saloner, 1986; Rotemberg and Woodford, 1992). In all these models, fiscal or monetary stimulus increases competition and output. In contrast, consumer-based models (Phelps and Winter, 1970; Afrouzi and Caloi, 2024) can also predict procyclical markups, and thus reduced competition during economic booms. A possible mechanism is that consumers become less price sensitive when the economy expands (Stroebel and Vavra, 2019). Taken together, the main drivers of competition and markups are changes in the competitive landscape and changes in consumer behavior. The relative importance of these channels will determine the aggregate cyclical pattern of markups.

The empirical literature on the cyclicality of markups is not in agreement. Nekarda and Ramey (2020), using aggregate time series evidence, find that markups may be procyclical. Others, e.g. (Chevalier and Scharfstein, 1995; Anderson et al., 2023) find countercyclical or acyclical markups.

In the absence of a consensus in the theory of markups or the empirical literature, we propose to construct a new time-varying measure of markups. Ideally, one would use estimates of marginal costs to estimate markups. Unfortunately, marginal costs are not directly observable and so it is difficult to construct a clean and time-varying measure of markups over a long sample period. There are two main ways in which one might proceed. One is using data on average costs, and one is by estimating demand. A prominent example of using average cost data is De Loecker et al. (2020). They document that, since the 1980s, there has been a long-run upward trend in markups, but they do not consider business cycle variation. In an alternative approach, DellaVigna and Gentzkow (2019) estimate demand elasticities. They do not allow for time variation in elasticities and therefore do not study cyclicality.[1]

In this paper, we use micro data from 2001 to 2022, which includes three recessions, to construct demand-side estimates of markups and attribute their dynamic variation into cyclical and trend factors. We proceed as follows: (i) We show that the demand elasticity at the store level is a sufficient statistic for markups that accounts for the competitive landscape and

[1]Indeed, De Loecker and Scott (2022) point out that, if properly implemented, the cost and demand side approaches will result in the same markups.

consumer factors. The markups can be computed using Lerner's index. We estimate revenue-weighted elasticities of demand at the store level by implementing the Hausman (1996) IV strategy and convert these to implied markups. Our estimates are at annual frequency, which allows us to study cyclicality. (ii) We build a dynamic panel data set of markups to identify trend and cyclical components of markup variation and use cross-sectional variation across regions to identify causes of markup variation. We find that markups indeed respond to both characteristics of the competitive landscape and consumer factors.

Following Gandhi et al. (2020), we use the first-order profit optimization condition for a grocery store to nonparametrically identify markups by associating the demand elasticity at the optimal price with the local average treatment effects (LATE) of log prices on log quantities (Imbens and Angrist, 1994). This general approach allows us to estimate markups when there are potentially nonlinear demand functions by calculating a local derivative close to the optimal price. It also allows us to abstract from the specific competitive structure. Instead, the store-level demand functions that we estimate can be interpreted as sufficient statistics for the variable of interest.[2]

We take as a starting point the possibility that elasticities may change over time. We assume that there may be a secular trend in elasticities, but also allow for higher-frequency cyclical variation. We use the food retail sector in the US as a setting to study the determinants of elasticities over time and across markets. There are several benefits of studying this sector. It is a large sector and for many households expenditure on food represents a significant fraction of discretionary spending (Cox and Harris-Lagoudakis, 2022) and therefore changes in food markups affect a large share of the population. Moreover, recent evidence suggests that consumers form their overall inflation expectations based on grocery bills (D'Acunto et al., 2021). The food retail sector also covers a wide variety of geographical locations, and there are many different goods, most of which consumers purchase frequently. Our data include the recent high inflation period during which food-at-home CPI inflation has outpaced headline CPI inflation. Food inflation has also been a high-profile source of lower consumer welfare contributing to the cost-of-living crisis.

We propose a two-step procedure to estimate time-varying elasticities and markups. In the first step, we use the well-known Hausman (1996) IV strategy to estimate own-price elasticities of demand.[3] We chose 25 large markets in the US in order to construct a geographically diverse sample of paired markets. The idea of Hausman is that one can use market pairs, for example New York City and Philadelphia, in order to identify cost shocks. We revenue-weight them in order for the estimated elasticity to reflect the typical consumer

[2]Gaarder et al. (2024) also estimate heterogeneous coefficient linear demand models with IV.

[3]The main other methods to estimate elasticities use demand control variables with high-frequency data (Levin et al., 2017; Brand, 2021) or impose covariance restrictions on supply and demand shocks (Döpper et al., 2022; MacKay and Miller, 2023). There have been alternative measures of IV variables such as production-side model-implied wholesale costs in De Loecker and Scott (2022).

basket. This approach leads to accurate market-category-year elasticities.

The scanner data we use comes from two sources—IRI (2001-2012) and NielsenIQ (2006-2022). Compared to the IRI data set, NielsenIQ covers both more goods and more stores in each market. For both data sets, there are sufficient observations to estimate demand elasticities for each market-category-year pair, and the estimation strategy produces precise and realistic results. Specifically, we find that (i) the Hausman IV strategy works—a test for weak instruments rejects the null for about 90% of elasticity estimations; (ii) estimates are precise under strong IV—more than 95% of their standard errors are below 0.5 and close to 99% of their t-statistics are above 1.96; and (iii) estimates are reasonable, i.e. almost all of them lie above one. Specifically, for only approximately 2% of the estimates we cannot reject the hypothesis of the elasticity being equal to one vs being above one at the 5% significance level.[4] We find large variation across markets, good categories, and, importantly, over time, validating our initial assumption of time-varying elasticities.[5]

In the second step, we pool elasticity estimates across markets and good categories to isolate variation in elasticities over time—both lower-frequency trend and higher-frequency cyclical variation. We find a pronounced downward trend in the elasticities. Using Lerner's transformation of elasticities, we convert our estimates of locally-linear demands to implied markups. The downward trend in elasticities thus implies an upward trend in markups. Using 2001 as the base year, our results imply that average markups—across all markets and good categories—went up 25% by 2019 and 101% by 2022.[6]

We also find evidence of procyclical markups. Our data extend from 2001 to 2022 and include two substantial contractions—the 2001 dot-com recession and the great recession of 2008, as well as a contraction of demand due to the tightening monetary policy in 2017. For all three of these contractions, we find large and statistically significant reductions in markups; the annual drops in markups lie between 8% and 22%. Our data set also includes the recession corresponding to the global pandemic year of 2020. However, as a result of aggressive monetary and fiscal stimuli and potential changes in preferences for online shopping, some aggregate demand measures increased during the pandemic and pushed markups up by 34%.[7] We find that, during contractions of aggregate demand, elasticities increase, while they decrease when aggregate demand expands. Markups are thus pro-cyclical.

Two related papers also consider annual elasticities. Brand (2021) considers a subset of nine food categories. Döpper et al. (2022) use the Berry et al. (1995) (BLP) method to estimate pooled US-level elasticities for 133 categories available in the NielsenIQ scanner

[4]This may be due to statistical errors given the expected 5% of false positives at this hypothesis testing.

[5]De Loecker and Scott (2016) also finds time-varying own-price elasticities for the beer market. Their paper combines production data and retail sector data.

[6]The average markup increased by 54% from 2006 to 2019, which is comparable to Döpper et al. (2022).

[7]See consumers' shifting to online shopping in Harris-Lagoudakis (2023). The risk of getting COVID may also make in-store consumers less price sensitive as they may stop shopping around to reduce that risk.

data, including many non-food categories. Neither paper uses an IV approach to estimate the elasticities. As a result, both may underestimate elasticities and overestimate markups. Also, neither paper analyzes the cyclicality of markups. Another difference from the existing literature is our focus on revenue-weighted elasticities and markups. We estimate that revenue-weighted markups, which results in a focus on those products that the typical consumer buys and leads to higher elasticities and lower markups.

We next proceed to explain the time variation in market-category-year markups. We use a structural dynamic panel to identify the factors, e.g. market-level and consumer factors, driving markups. This analysis allows us to measure the effect of increases in aggregate demand, including those resulting from policy, on markups while controlling for other factors. This regression framework also lets us identify conditional cyclicality, i.e. measure the conditional contribution of various drivers of markups. We use a dynamic partial adjustment model (Balestra and Nerlove, 1966) to estimate the effects of different drivers on markups. In a dynamic panel, we include four sets of variables: demographic trends, market concentration, aggregate demand, and real interest rates. Then we treat these factors as exogenous and construct a structural historical decomposition of the markups panel in a manner similar to the SVAR literature (Kilian and Lütkepohl, 2017). This structural dynamic panel provides estimates of effects of market-specific and aggregate factors from time-varying cross-sectional heterogeneity in markups.

When explaining variation in markups we use measures capturing both factors affecting demand and those affecting market structure. We find several effects. First, we identify a positive effect of household income, measured by per capita GDP. When people's income declines, for example, because of a more severe recession than experienced by other markets, markups decrease as demand elasticities go up. Aguiar et al. (2013) also find that consumers become more price sensitive in response to income losses, e.g. by shopping around more.

Second, an increase in population leads to lower elasticities and therefore higher markups. An increase in population might be a sign of improved economic performance or anticipated income growth, potentially making consumers less price sensitive.

Third, higher housing prices result in higher markups in both the current year and the long-run. The positive effect may come from the wealth effect for home owners (Stroebel and Vavra, 2019).

Fourth, we also find effects of market factors, such as store concentration (measured by the number of stores per capita) and product concentration or the breadth of product choices within each good category (measured by the number of UPCs per category). These market factors, especially the reduced product concentration, have put strong downward pressure on markups over the sample period. Although higher retailer concentration significantly increases markups, its effect size tends to be secondary.

Fifth, we find a strong effect of the real interest rate. Lower real rates, e.g. negative

rates, lead to higher markups. This suggests a direct effect of inter-temporal substitution on aggregate demand. Importantly, this effect is present after controlling for current household incomes, proxied by per capita GDP. Our finding highlights an important transmission mechanism of monetary policy that acts through changes in optimal markups. What this means is that monetary policy has a direct impact on prices. This effect is in addition to indirect effects operating through the income channel.

We use our regression results to do a structural historical decomposition into the four major groups of factors that we have identified. We find that these factors can explain variation both over time and across markets. The major driver of markup growth over the sample was an increase in aggregate demand, as proxied by our set of macro factors. In contrast, market-concentration measures put downward pressure on markups and population growth had a significant impact only on a few markets.

The real interest rate, which we include separately from the macro factors, put important supportive pressure on markups during the zero-interest-rate period of 2008 to 2016. This effect is separate from the effect of aggregate demand, which, during this period, put large downward pressure on markups. As such, one can interpret the accommodative monetary policy as stabilizing prices.

The remainder of the paper is organized as follows. Section 2 introduces our methodological approach to elasticity and markup estimation. We also discuss the theoretical IV approach to the elasticity estimation. Section 3 discusses the retail scanner data that we use, outlines some of the data choices we make, and presents estimation results. It also discusses trend and cyclical variation in elasticities and markups. Section 4 analyzes factors that drive markup dynamics, and their historical contributions. Section 5 discusses the implications of these results. Section 6 concludes the paper.

2 Nonparametric identification of markups

2.1 Store price-setting model

We start by introducing a model that is sufficiently flexible to allow us to estimate time-varying demand elasticities and markups in a market with many products and competing stores.

Let $Q_{si\tau}(P_{si\tau}, P_{-si\tau}, X_{si\tau})$ be a smooth downward-sloping store-level demand curve faced by store s for product i at time τ. It depends on the price charged by competitors $P_{-si\tau}$ and the demographic characteristics of its consumers $X_{si\tau}$, which will depend on the store's location. Note that stores may charge different prices since they are in different geographic locations. Suppose that all stores follow a Bertrand price competition model. In the Bertrand-Nash equilibrium, competitor stores set equilibrium price $P^*_{-si\tau}$, which in general will be

different across stores. We assume constant unit costs equal to $c_{si\tau}$. A given store s then sets its own price by maximizing profits according to:[8]

$$\max_{P_{si\tau}}(P_{si\tau} - c_{si\tau})Q_{si\tau}(P_{si\tau}, P^*_{-si\tau}, X_{si\tau}). \tag{1}$$

The first order condition for the optimal price $P^*_{si\tau}$ implies

$$(P^*_{si\tau} - c_{si\tau})\frac{\partial Q_{si\tau}(P^*_{si\tau}, P^*_{-si\tau}, X_{si\tau})}{\partial P_{si\tau}}|_{P_{si\tau}=P^*_{si\tau}} + Q_{si\tau}(P^*_{si\tau}, P^*_{-si\tau}, X_{si\tau}) = 0. \tag{2}$$

Markups $\mu_{si\tau}$ are then set according to:

$$1 + \mu_{si\tau} = \frac{P^*_{si\tau}}{c_{si\tau}} = \frac{e_{si\tau}(P^*_{si\tau}, P^*_{-si\tau}, X_{si\tau})}{e_{si\tau}(P^*_{si\tau}, P^*_{-si\tau}, X_{si\tau}) - 1}, \tag{3}$$

where $e_{si\tau}(P^*_{si\tau}, P^*_{-si\tau}, X_{si\tau}) = -\frac{\partial \log Q_{si\tau}(P_{si\tau}, P^*_{-si\tau}, X_{si\tau})}{\partial \log P_{si\tau}}|_{P_{si\tau}=P^*_{si\tau}}$ is the absolute elasticity of residual demand that store s faces at the profit maximizing price in the Bertrand-Nash equilibrium.

Assuming that, in the data, we always observe equilibrium prices in the neighborhood of $P^*_{si\tau}$ for all stores s, one can compute the markups $\mu_{si\tau}$ using a nonparametric estimate of the local average treatment effect (LATE) of the logarithm of the own price on the logarithm of residual demand at every store. Therefore this LATE parameter is a sufficient statistic for the optimal markup charged by the store. A similar LATE approach was recently introduced by Munro et al. (2023) to study the impact of egg subsidies in rural India. Similar to us, they do not require a structural parametric model as long as the observed prices remain close to market equilibrium. However, unlike them, we work directly with differentiated products and Bertrand-Nash competition. This setup is also related to Gallego et al. (2019), who consider optimal price setting theory for a monopolist with a general smooth demand function, so that it can be approximated with a linear demand function in the neighborhood of the optimal price.

If we estimate a log-log demand specification of a locally linear model, the markups formula holds under the Bertrand-Nash competition of profit maximizing firms with a general and time-varying residual demand curve. Thus, this result can be seen as a nonparametric identification of markups using the first-order condition as was proposed in Gandhi et al. (2020) for identification of the production functions.

The formula (3) also applies to a parametric demand function. Originally Cournot (1838) derived the optimal markup for a monopolist. Later, Lerner (1934) suggested to use this formula for $\mu_{si\tau}$ as a measure of market power, hence this index is widely referred to as

[8]We could include fixed costs in this formula, but omit them because they do not affect the maximization problem.

Lerner's index. The index has been used extensively in recent work under the assumption of a constant elasticity of demand (for example, DellaVigna and Gentzkow, 2019).

2.2 Panel-IV regression approach to demand elasticity estimation

In order to take our conceptual model to the data, we provide a mapping from the model variables to the variables in the data.

In the data, which we discuss in more detail in the next section, we observe multiple UPC (barcode) varieties i of good category j, for example, different types of cereal. We assume that within a year the actual price $P_{si\tau}$ fluctuates in the neighborhood of the optimal price $P^*_{si\tau}$. This price variation can be driven, for example, by promotions imposed by wholesale suppliers.

The most granular approach would be to calculate store-level elasticities for each product variety i in a given week τ. This is the level at which we observe the data, and so is trivially impossible. We therefore must aggregate the data in one or more dimensions to be able to calculate elasticity estimates. Moreover, to achieve sufficient power, we need to aggregate further and calculate an average elasticity for a specific group of observations. Naturally, by aggregating across varieties and stores, higher power comes at some expense of losing resolution. This means that we need to choose a sample over which we are averaging and then choose weights in order to calculate the within-sample average. We next discuss grouping size and the associated estimating equation; after that we turn to the weighting scheme.

Since we are interested in cyclical variation at the national level, we choose to estimate average demand elasticities at the market-good-year level. That is, to estimate the slope, we assume that it is approximately constant across weeks τ in a year t, across stores s in a region r, and across varieties i (one type of cereal) in a good category j (any type of cereal). That is, for example, we estimate the average elasticity of milk in Sacramento in 2020.[9] These estimates should be understood as averages of the underlying demand derivatives (elasticities) in the neighborhoods of the respective optimal prices in a given year for a given geographical location.

In our model, competitor prices and consumer characteristics can affect demand. Having pooled the data, we now need to capture variation in these other demand shifters that may vary across goods, across stores, and over time. That is, a family-sized cereal box may have different demand than a smaller size; demand for candy may be different in the weeks leading up to Halloween; and demand in big-box retailers may be different than for a small convenience store. We use a saturated regression model (see, for example, Abadie

[9] Alternatively, one may further subdivide good categories into smaller groups like organic and conventional good, as was done in the ongoing work Gafarov et al. (2024). Other options are to divide markets r into store chains (DellaVigna and Gentzkow, 2019), divide big categories into subcategories by brand or food labels, or work with a quarterly (or monthly) resolution instead of the annual one we choose.

et al., 2020; Bugni et al., 2023) to control for the majority of demand factors $X_{si\tau}$ and the prices of substitutes using group fixed effects $\{upc_{rit}, store_{sjt}, week_{rj\tau}\}$. The idea is that a sufficiently large set of fixed effects controls for market conditions and consumer income nonparametrically, which is more flexible and potentially more robust than including those controls directly and assuming a specific functional form.

$$\log Q_{si\tau}(P_{si\tau}, P_{-si\tau}, X_{si\tau}) = -e_{rjt} \log(P_{si\tau}) + upc_{rit} + store_{sjt} + week_{rj\tau} + \varepsilon_{si\tau}. \tag{4}$$

On the LHS, we see that the resolution of our data is at the store-UPC-week level with subscript $si\tau$. On the RHS, we note that the elasticity e_{rjt} varies across regions r, good categories j and years t, which is our chosen level of aggregation. It strikes a balance between being sufficiently specific to capture important differences between goods, across regions, and, importantly, over time. At the same time, this level of aggregation delivers enough power to be able to estimate elasticities with a relatively high degree of accuracy.

Three sets of fixed effects capture various demand effects: We assume that demand depends on (i) the UPC, i.e. the specific good, and that this effect varies across regions r and over time t, (ii) the store, an effect that varies across good categories j and time t, and (iii) the week τ, an effect that varies across regions r and good categories j. Our fixed effects capture differences in demand across UPCs, e.g. due to quality, across stores, e.g. due to location, and across time, e.g. due to seasonal variation in demand.

We assume that prices of competitors enter the demand function mostly through a common market-level demand shock. In other words, demand at a specific store reacts primarily to changes in the average regional price, so changes in specific competitor prices affect demand mainly through the average. We can therefore capture such competitor-driven demand changes, $P_{-si\tau}$, together with the consumer-side demand shocks, $X_{si\tau}$ using weekly fixed effects .

There is a recent literature concerned with automatic selection of groupings for time-varying fixed effects, for example, Bonhomme and Manresa (2015); Chetverikov and Manresa (2022). These procedures are computationally prohibitive for large datasets such as ours.[10] Such automatic procedures could be more promising when one is estimating demand for one individual product category as opposed to the entire national market, which is the focus of our paper.

We now discuss the weighting scheme when estimating the weighted average elasticity e_{rjt} for a specific market, food category, and year (r, j, t) in equation (4).[11] Our goal is to estimate representative markups of products that a typical consumer buys in a typical store

[10] An estimation of approximately 30,000 elasticities using regressions (4) takes about two weeks even if group three-way fixed effects are pre-specified and not chosen optimally.

[11] For discussion of large sample limits of OLS and IV estimates in the presence of heterogeneous observations see, for example, Section 3.2 in White (2014).

in a typical week.

We use the year-round revenue share of a given UPC in a given market to estimate revenue-weighted least squares average elasticity e_{rjt}.[12] We then use these elasticities to calculate optimal markups. It is useful to think about implications and potential advantages of revenue-weighting the data rather than using an equally-weighted approach. There are two important differences relative to equally-weighting. First, sales tend to be concentrated in a small share of UPCs. This means that an equally-weighted approach will effectively overweight UPCs that have low sales. More specialized products tend to be purchased by dedicated consumers, may have lower elasticities and lower revenue shares. For example, some customers may have inelastic demand for highly specialized products, e.g. specific organic products or fair trade labels. Thus, an equally-weighted estimation approach results in lower average elasticities than the revenue-weighted alternative. We discuss results from such an approach in the appendix.[13] Second, due to limited attention of a store, the UPCs that bring the most revenues are more likely to be priced optimally and thus reflect our theoretical model of profit maximization.

2.3 Hausman's IV for cost shocks

The remaining variation in demand shocks not captured by the fixed effects is included in the residual term $\varepsilon_{si\tau}$. However, shocks to demand may be correlated with price changes since they are determined together with product quantities in market equilibrium (Working, 1927). This correlation may therefore cause endogeneity bias of elasticities towards zero.

The standard solution to the endogeneity problem when estimating demand elasticities is to use an instrumental variable that moves together with the supply shock, but is orthogonal to any remaining demand shocks in equation 4. To address this endogeneity problem, we follow Hausman (1996) and Nevo (2001), and instrument the log of a store-specific weekly product price in a market of interest by the quantity-weighted average of log weekly prices of the same product sold at all stores in the paired market(s), i.e. markets that are geographically close.

We group the data into major markets, which are defined in our data (see details in the next section). We match neighboring markets and construct market pairs. Most products sold in neighboring markets (but not distant markets) are the same. This enables us to instrument product prices in a market of interest by those in its neighboring market(s). We select 12 pairs of 25 largest markets that are from different US regions, thus covering the East, South, West and Midwest of the country (see Figure 1). We believe that by focusing on geographically close markets we better capture common local cost shocks such as labor

[12] See Appendix for further details on the estimation strategy.

[13] Estimates are reported in Table B.6. The revenue-weighted estimates are reported in Table 1.

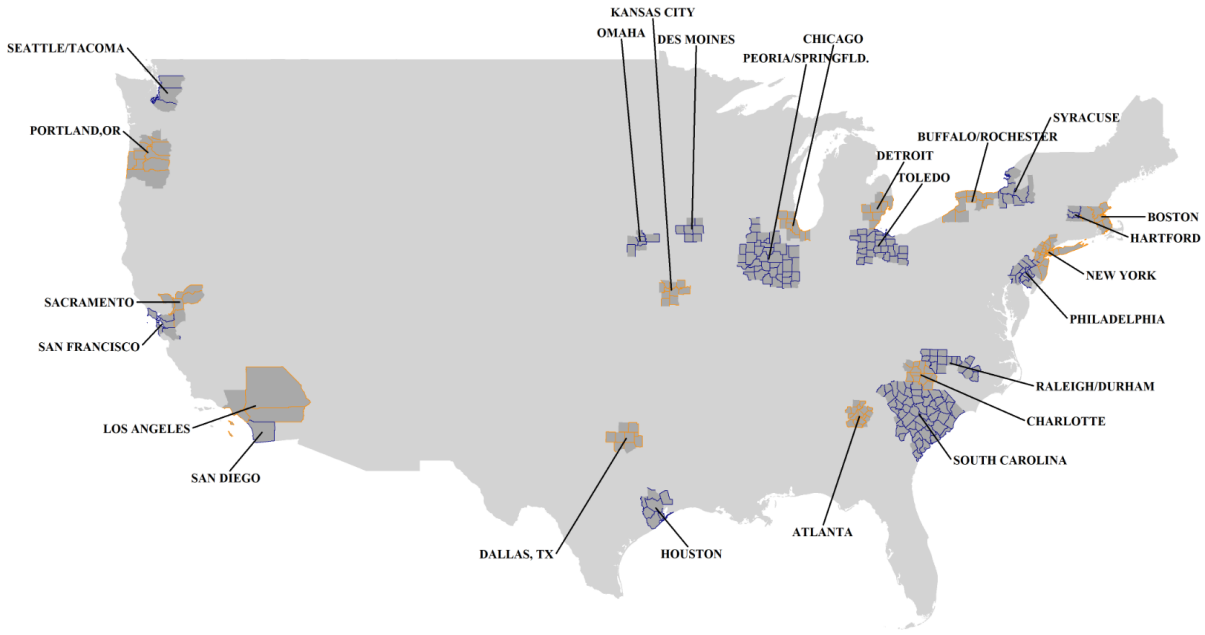


Figure 1: The 25 Paired Markets.

Note: Paired markets have county borders colored in blue and orange. See specific pairs in Table B.1.

and transportation costs.[14]

Note that product UPCs are manufacturer-specific, i.e. a product with the same UPC sold in the paired market comes from the same manufacturer, which delivers the relevance of our price instrument. For example, suppose that a wholesale supplier introduces promotional coupons that temporarily lower the price of a particular UPC of a cereal in a given week. These coupons would be passed to all the sellers of this product and constitute a positive supply shock. The coupon then will lead to a price decline for that UPC relative to the whole product category as required for the estimation of the own-price elasticity of demand. In contrast, if producer input costs change, it will impact the whole good category and will be captured by the weekly fixed effect together with category-level demand shocks. Of course, there may be idiosyncratic good-specific demand shocks in geographically close areas resulting in changed prices that are not driven by supply shocks. For example, locally consumers may stop purchasing a specific type of lettuce because of a reported e-coli outbreak. Stores may reduce prices for that UPC in response to such a demand shock. Since we estimate an average of price elasticity of hundreds of UPCs in a market, e_{rjt}, such an occurrence would not materially affect the final estimated elasticity unless it affects the majority of the UPCs in the category. Indeed, as discussed in Section 3.2 below, we find that this Hausman-type price IV is strong for most of our demand elasticity estimates, which suggests that our strategy

[14]Our implementation of the Hausman IV strategy is different from DellaVigna and Gentzkow (2019) who used the average price for a given UPC across the entire nationwide store chain.

works.

3 Estimation of Markups

3.1 Retail scanner data

Our aim is to estimate markups in order to study their dynamics over time as well as heterogeneity across markets. As we have shown above, we can estimate the LATE of product prices on demand and, using that, construct estimates of markups. For this, we require estimates of demand elasticities, which are estimated using data on prices and quantities for various food categories, markets and years.

We use retail scanner data from the IRI Marketing Dataset (Bronnenberg et al., 2008) and the NielsenIQ Dataset.[15] The IRI retail scanner data covers 12 years, from 2001 to 2012 while the NielsenIQ scanner data span 17 years, from 2006 to 2022. They both have weekly transaction records of products sold by physical retail stores across markets. We focus on panel food stores that have reported transactions across years in the IRI and NielsenIQ datasets because we study time variation of markups. In the following, we describe the key features of the data that are relevant to our empirical estimation strategy.

Markets. As discussed, we choose 12 pairs from 25 markets (Figure 1).[16] There are 50 IRI markets in total. The other 25 markets, which are not used in this paper, are mostly small or do not have clear neighboring markets for pairing. An IRI market consists of one or multiple adjacent counties; we match NielsenIQ grocery store locations to find the IRI markets to which they belong.

Markets vary significantly in size. The largest markets are New York and Los Angeles. In total, all markets had 123.8 million residents in 2010, more than 40% of the 2010 US population. Taken together panel stores in the 25 markets represent over 33% of the observations in the IRI scanner data, and contain a large but smaller proportion of observations in the NielsenIQ scanner data. Table B.2 shows a list of all markets, their population, number of observations and number of stores. Overall, these markets are broadly representative.

Food categories. For estimation purposes we group UPC products into 16 food categories defined in the IRI sample. In the NielsenIQ sample, we estimate elasticities for 58 food categories, defined by using the NielsenIQ food groups. We include the complete list of food categories in Table B.4 and discuss additional implementation details in Appendix A

[15] We obtain the NielsenIQ Retail Scanner Dataset from the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.

[16] We combine Omaha and DeMoines and pair them with Kansas City. Table B.1 outlines market pairs.

We estimate demand elasticities at the food category level rather than at the more disaggregated UPC level. As shown in Table B.5, markets do not all sell the same goods in terms of UPC. For example, some markets sell more and some sell fewer goods. Nevertheless, goods sold in neighboring markets are mostly the same. This allows us to instrument good prices in a market of interest by those in neighboring market(s) when estimating demand elasticities.

3.2 Estimates of market-category-year elasticities

We are now ready to estimate elasticities at the market-category-year level. When implementing our estimation strategy we find the proposed Hausman-type IV works for most weekly product prices at the UPC level across markets and years. This is due to the fact that the selection of products sold in neighboring markets is almost the same each year. In total, we successfully match 90.8% and 91.2% of price observations with their IVs in the IRI and NielsenIQ samples, respectively. These market-category-year specific IV regressions deliver us 29,195 raw demand elasticity estimates.

We find that the price instrument is relevant (strong) for the vast majority of the market-category-year combinations. Following the rule of thumb of (Staiger and Stock, 1997), we drop estimates with an F-stat below 10, which represents about 10.5% of the data, with a relatively higher percentage within the IRI sample possibly due to its smaller number of observations for each elasticity estimation. As shown in Section 2.1, we measure the absolute demand elasticity to capture the price responsiveness. Among the strong-IV demand elasticity estimates, less than 1% have an incorrect sign, most likely due to estimation errors in a typical IV regression; we drop these estimates as well.

Additionally, because of finite samples, outliers can have an outsized effect on sample moments. Therefore, sample means and least squares estimates, both of which we calculate, can be affected by outliers. We drop the top and bottom 1% of the remaining elasticity estimates each year. Our final sample has 25,367 demand elasticity estimates with 14.3% in the IRI sample and 85.7% in the NielsenIQ sample.

In order to assess the validity of the price instrument, we compare OLS (equally weighted), WLS and IV (revenue-weighted) estimates for the final sample of strong instruments. Figure 2 reports the estimated kernel densities for the three elasticity distributions. There is a strong shift in the distribution from OLS to WLS, suggesting that there are a lot of UPCs with fairly low market shares and appear to not be representative of what people purchase. One reason could be that more differentiated, niche products with low market shares are less elastic. We next note that, the strong shift from WLS to weighted IV indicates validity. One way to get a sense of the magnitude is to compare medians; the OLS median is about 1.16, the WLS median is 1.65 while it is about 2.22 for the weighted IV estimates. Also,

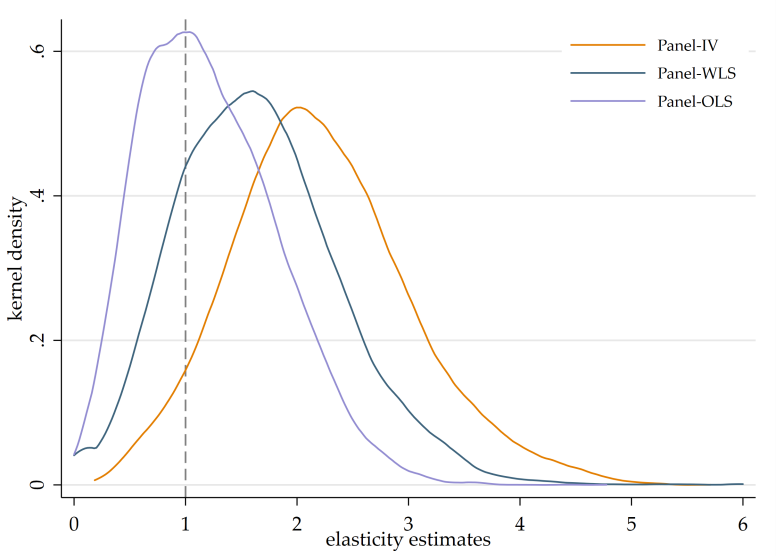


Figure 2: Smoothed density estimates of the distributions of OLS, WLS, and IV demand elasticity estimates.

the fraction of estimates below one is equal to 40.4% for the OLS, 19.5% for WLS but only 5.6% for the IV estimates. We conclude that our instrument corrects for a significant part of the downward endogeneity bias, while the revenue weighting adjusts for the composition of consumers' shopping patterns.

Another way to show the validity of our approach is to compare our elasticity estimates with estimates from the literature. We note that the comparison is not exact since we are estimating a revenue-weighted average elasticity and so place higher weight on potentially higher-revenue products. We therefore may expect our estimates to be slightly higher than those obtained using BLP. Cereal is commonly studied in the literature and therefore an appropriate category for comparison. For the IRI sample, our mean elasticity estimate is about 2.8 for the Los Angeles market during 2007-2010. For the same market and time period, Richards and Hamilton (2015) obtain an elasticity mean of 2.2 based on a BLP model, which is comparable with our mean elasticity estimate. For the NielsenIQ sample, our median elasticity estimate is about 2.7 for all 25 markets during 2007-2017. For the same time period, Backus et al. (2021) have almost the same median elasticity estimate, although they include fewer diverse markets in their sample of interest. In sum, our cereal elasticity estimates are comparable to those obtained using BLP for both the IRI and NielsenIQ datasets.

We focus on an elasticity cutoff of 1 because optimal price setting, according to equation (3), implies that if we observe an estimated (absolute) demand elasticity below 1, the observed price will not have been set based solely on profit maximization for the specific food category. The fact that we find a much smaller percentage of below-one estimates for the IV strategy indicates that this estimation method is more precise.

Table 1 below reports summary statistics for the final sample. We report statistics separately for the IRI and NielsenIQ samples. We note that the average demand elasticity estimate in the IRI sample is higher than that in the NielsenIQ sample. This pattern is mainly due to the downward trend in demand elasticities, which is one of the main findings of our paper and which we start to discuss in more detail in subsection 3.3.

We compute standard errors clustered at the store and week levels in two ways to allow for arbitrary correlations of the error term $\varepsilon_{si\tau}$ across products within each store and across stores within each week. The standard errors are small, implying that the elasticities are estimated precisely. Close to 99% of the t statistics are above 1.96. Indeed, 90% of the t-statistics are above 7.5 (IRI) and 5.7 (NielsenIQ). This is because we have sufficient observations and strong price IVs in nearly every elasticity regression, as reported in Table 1.

Table 1: Summary Statistics of Demand Elasticity Estimates.

	IRI sample[a]			NielsenIQ sample[b]		
	mean	p10	p90	mean	p10	p90
point estimate	2.92	1.99	3.90	2.10[c]	1.17	3.07
standard error	0.23	0.10	0.40	0.17	0.06	0.34
t statistics	16.85	7.49	27.81	18.49	5.69	33.99
obs in an estimation (thousands)	123.3	19.5	263.6	1155.4	66.8	2906.4
Cragg-Donald F statistics	88.5	14.3	180.2	1335.0	31.0	2059.7

Note: [a]The IRI sample has 3,619 elasticity estimates. [b]The NielsenIQ sample has 21,748 elasticity estimates. These samples include elasticity estimates that are obtained under the strong price IV and exclude those that have negative values. For each sample, we also trim off the upper and lower 1% of the remaining elasticity estimates in each year. See detailed descriptions in the text above. [c]The NielsenIQ sample has lower elasticity estimates than the IRI sample. The downward trend in demand elasticities over time, as explained in Section 3.3, largely contributes to the lower elasticity estimates in the NielsenIQ sample of more recent years. The NielsenIQ sample also has a small proportion of price-inelastic goods such as baby food, spices, seasoning, extracts and eggs.

As noted in the discussion of Figure 2, 5.6% of IV estimates are below one. Less than half, 2.4%, are significantly below 1 when using the conventional 5% significance level. In terms of revenue shares, these numbers are 3.6% and 1.6%, respectively.

There are several reasons for this pattern. First, there is a finite sample estimation error and so in any distribution we expect to find some percentage of estimates that are significantly below one. Second, there is variation in the ability of the instrument to correct for the endogeneity bias that is present when using OLS. Therefore, for some market-category combinations, the correction is smaller in magnitude and therefore results in estimates below one. Third, in the latter part of the sample, during and immediately after the pandemic, there was higher uncertainty in demand behavior and marginal costs. This may have made it more difficult for stores to optimally update markups in a timely manner. It is therefore possible that for some of the products in specific markets, stores may have used price setting

behavior based on previous rather than current information. We discuss implications of these potential factors resulting in estimates below one further in later sections.

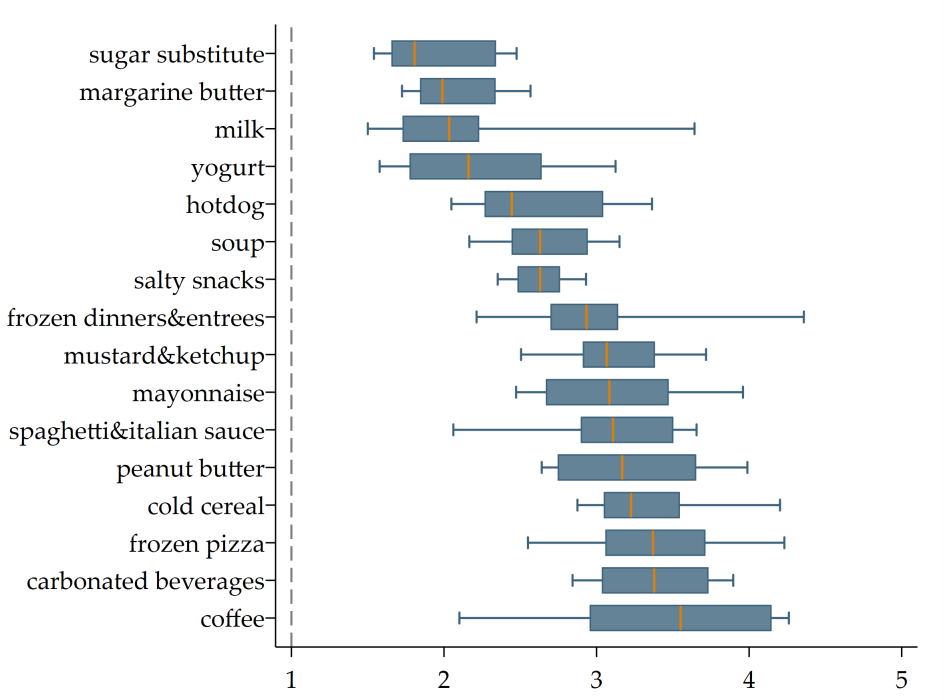


Figure 3: IRI Demand Elasticity Estimates by Food Categories in 2010.

Note: This boxplot shows the elasticity distribution across 25 markets for the 16 food categories in the 2010 IRI sample. We order food categories by their median elasticity estimates colored in dark orange. Each boxplot consists of two parts—one box and two whiskers. A box ranges from the lower quartile to the upper quartile of elasticity estimates with the median in the middle. The two whisker ends are designed to represent for the 10 and 90 percentiles of the elasticity distribution, respectively.

Our demand elasticity estimates also exhibit notable heterogeneity across markets and goods, which is in line with the empirical literature (DellaVigna and Gentzkow, 2019). For example, Figures 3 and 4 show that IRI demand elasticity estimates generally range from one to five across food categories and markets in 2010. Within each food category (market), these elasticity estimates also have sizable differences across markets (food categories). NielsenIQ demand elasticity estimates have similar features, as shown by Figures B.1, B.2, and B.3 in the appendix. We will tease out these notable level differences among market-category-year specific demand elasticity estimates through market- and category-specific fixed effects in our factor analyses for elasticity and markup dynamics in Section 4.

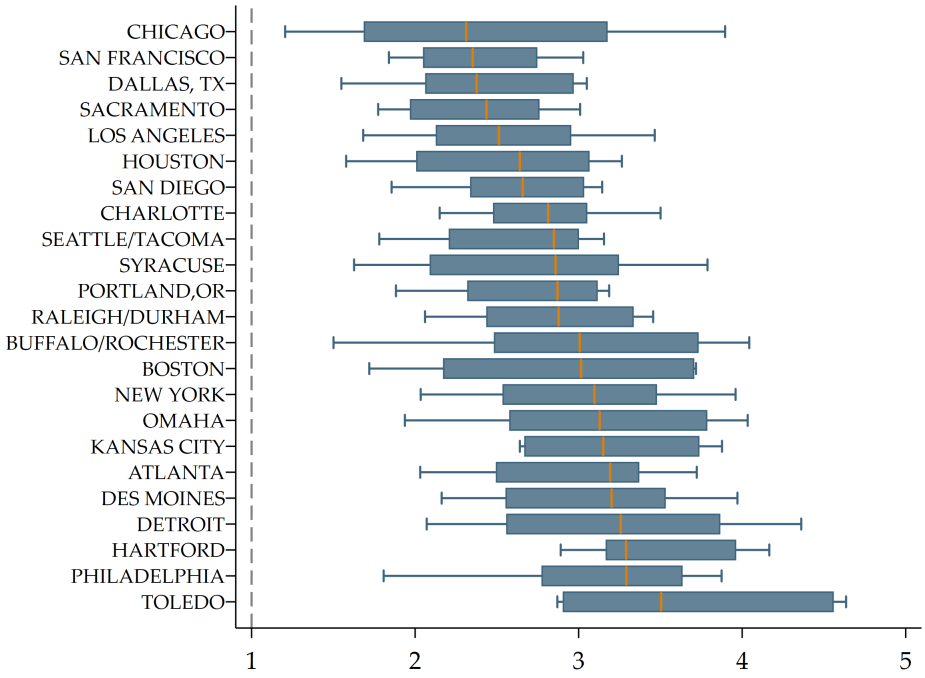


Figure 4: IRI Demand Elasticity Estimates by Markets in 2010.

Note: This boxplot shows the elasticity distribution across 16 food categories for 23 markets in the 2010 IRI sample. (South Carolina and Peoria/Springfield markets have limited elasticity estimates in this particular sample so that we do not show their elasticity distributions here.) We order food categories by their median elasticity estimates colored in dark orange.

3.3 Cyclical and trend variation: elasticities

We now proceed to analyze time variation in average elasticities by pooling the IRI and NielsenIQ market-category-year estimates. Next, we discuss markup implications for the food retail sector in Section 3.4.

We begin by summarizing changes in the elasticity estimates over time. Each year, we calculate the mean, 25th and 75th percentiles of the full distribution of estimated market-category elasticities. Figure 5 reports these three time series.[17] There are three patterns that we immediately notice. First, there is a pronounced downward trend in demand elasticities over the sample period. Mean elasticities decline from 2.55 (peak in 2004) to 1.55 (in 2022).

Second, at times of contraction in aggregate demand, elasticities increase, implying counter-

[17]Due to differences in store attributes and good coverage, the estimates of elasticities across the two data sets are not directly comparable. Indeed, we find that the average elasticity estimates during the overlapping part of the sample is not exactly the same. The IRI data set is much smaller in scope (fewer goods, stores, and years); we therefore shift IRI elasticity estimates by a constant to match that estimated in the NielsenIQ dataset. One reason for the difference may be the fact that a NielsenIQ food category has relatively more substitutable products at the lower UPC level, as shown in Table B.5.

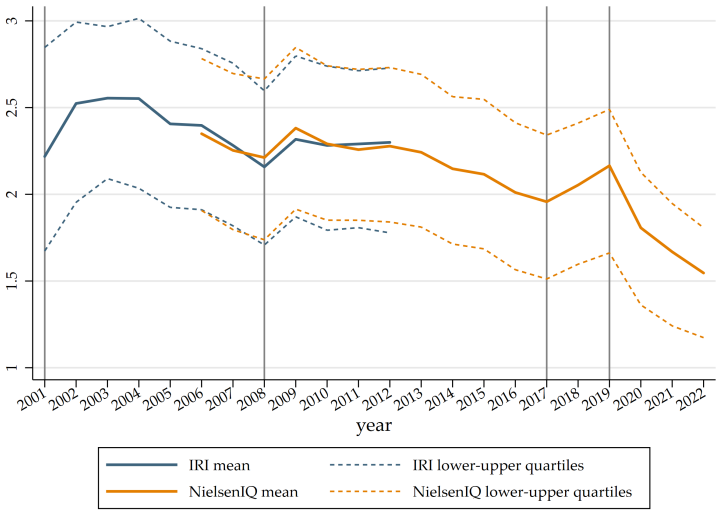


Figure 5: Demand Elasticity Estimates over Time.

Note: For each year, we report mean, upper, and lower quartiles of the pooled market-category distribution of elasticities. The four vertical gray lines refer to important aggregate demand shocks—the recessions in 2001, 2008-2009, the monetary contraction in 2017 and the COVID pandemic and associated recession in 2019-2020. As explained in the text, due to differences in store attributes and good coverage between the IRI and NielsenIQ samples, we shift IRI demand elasticity estimates down by a constant so that they have the same average elasticity estimate as NielsenIQ's during the overlapping period of 2006-2012.

cyclical elasticities.[18] For ease of identifying important aggregate demand shocks, specifically recessions in 2001, 2008-2009, monetary contraction in 2017 and the COVID-19 pandemic and associated recession in 2019-2020, we add vertical lines to Figure 5 to denote the first year of these events. For ease of discussion, we are highlighting these events now; later, as part of the factor decomposition in Section 4, we discuss them in more detail.

We check that the cyclical and trend variation is present both across good categories and markets (see the examples shown by Figures B.5 and B.6). The pattern is thus not driven by potential time-varying changes in the composition of the sample. This motivates us to pool the IRI and NielsenIQ elasticity estimates into one sample for our main analysis below.

Third, we note that the dynamic pattern of elasticities is present not only in the mean elasticity time series but is also reflected by the entire distribution. This pattern is reflected by the evolution of the quartiles of the elasticity distribution.[19]

In addition to a visual inspection of the estimates in Figure 5 it is useful to decompose these changes into a secular trend component and a higher frequency cyclical component. The former part captures long-run changes in consumers' preferences and shopping behavior as well as market structure. The latter part captures the fluctuations of consumers' purchasing

[18] We note that the increase in elasticities in 2008-2009 is present both in the IRI and NielsenIQ data.

[19] Recently De Loecker et al. (2020) argue that the upward trend in markups across US industries is driven, to a large extent, by the upper tails of the markups distribution. The results in Figure 5 suggest declining elasticities and therefore increasing markups across all quantiles of the distribution.

power over business cycles.

We run WLS of the market-category-year demand elasticity estimates in the pooled sample on year dummies and a linear year trend, using within-year market-category revenue shares as weights. We use market-category-specific fixed effects to control for level differences in demand elasticity estimates across markets, categories, and samples. In addition, we drop year dummies for 2004 and 2019, the two reference years, such that the cyclical variation around the linear trend has a statistically zero sum for the years 2001-2019.[20]

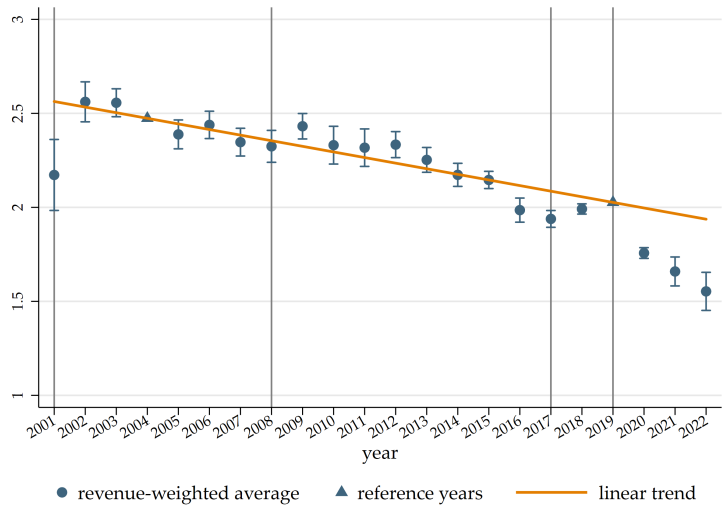


Figure 6: Trend and Cyclical Variation of Demand Elasticity Estimates.

Note: This figure shows trend and cyclical variation of demand elasticity estimates in the pooled IRI and NielsenIQ sample, where we shift IRI estimates to match the NielsenIQ average during the overlap period (same as Figure 5). Details of our estimation are discussed in the text. The four vertical gray lines indicate aggregate demand shocks. The spikes show the 95% confidence intervals of the revenue-weighted average annual demand elasticities relative to the linear trend. Standard errors are clustered at the market level.

Figure 6 provides a graphical representation of the trend and cyclical variation over time. We find a substantial and highly statistically significant downward trend in demand elasticities of 0.03 per year, and a sizable increase of, on average, 0.21 during years of aggregate demand contractions. We note that cyclical changes in demand elasticities are large during contractionary aggregate demand shocks. For example, elasticities increase by 0.42 from 2001 to 2002, during the recession associated with the dot-com bubble bursting, and 0.14 after the 2008 financial crisis. Given these economically large changes over time, we next discuss their implications for markups.

[20]The pandemic and post pandemic years are not included in the trend calculation since we find evidence that there was a pandemic shock that resulted in lower elasticities. We discuss this in more detail in Section 4.

3.4 Cyclical and trend variation: markups

We now explore the implications for markups. Following our discussion of the Bertrand-Nash equilibrium price setting in Section 2.1, we calculate markups, $\mu_{rjt} = 1/(e_{rjt} - 1)$, implied by the local elasticity of demand estimates according to equation (3).

For each market-category-year we calculate implied markups, delivering a panel data set of markups. Figure 7 plots pooled sample average markups as well as 25th and 75th percentiles over time.[21]

From the distribution of elasticities in Figure 2 we know that some estimates lie below one, though only a small fraction of estimates are smaller and statistically different from one. For these elasticity estimates, the markup formula cannot be applied. We hypothesize that stores may not be setting prices optimally, which in turn implies that the markup formula will not measure markups. We therefore adjust these estimates in the following way. When a market-category-year-specific elasticity estimate is below 1.2, we replace the associated markup with the most recent one for which the elasticity is higher than this cutoff (so that the markup lies below 5). The idea is that the store may have kept the previous markup constant even though consumer behavior and thus the estimated elasticity have changed. There are about 9.9% of elasticity estimates that are below 1.2. Among them, there are market-category-specific elasticities that are below 1.2 over all years, which contributes to about 0.9% of our elasticity estimates. For these elasticities, we can not do the replacement and thus do not calculate their corresponding markups. Thus, we apply the replacement to the remaining 9.0% of elasticity estimates.[22]

Figure 7 shows that there is a strong upward trend in markups over the sample period. The cumulative change over the sample period is about 100%. However, the annual growth rate has not been constant. There are noticeable and statistically significant decreases in markups in particular at times of contractionary aggregate demand shocks, mirroring the decreases in elasticities (Figure 5).

Markups are a non-linear transformation of elasticities and so we next check that the cyclical and trend components of changes in elasticities are reflected in the markup dynamics. Following our estimation for elasticities (Figure 6), we run WLS of the market-category-year demand markups on year dummies and a linear year trend, using within-year market-category revenue shares as weights.

Figure 8 shows upward trend in markups through 2019 as well as the acceleration in markup increases from 2019 until 2022. We can also see the cyclical component—in response

[21] As before, we shift IRI markups by a constant to match NielsenIQ markups in the overlap period.

[22] For the market-category-specific elasticities that have estimates above 1.2, the replacement applies to about 5.7% of the data before 2020 and always below 10% over the years. From 2020 to 2022, the replacement fractions are 16.0%, 21.3%, and 27.1% respectively.

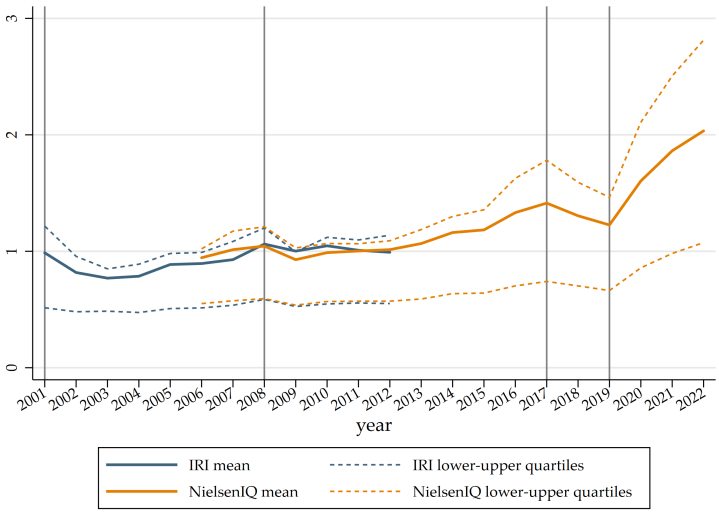


Figure 7: Implied Markups over Time.

Note: We calculate markups from demand elasticity estimates based on equation (3) in the main text. When a market-category-year-specific elasticity estimate is below 1.2, we replace the associated markup with the most recent one for which the elasticity is higher than this cutoff (so that the markup lies below 5). This replacement applies to about 9.0% of the data. We drop about 0.9% of the data for which market-category-specific elasticities are below 1.2 over all years. The four vertical gray lines refer to important aggregate demand shocks. Similar to Figure 5, we shift IRI markups by a constant to match the NielsenIQ markup average during the overlap period.

to aggregate demand contractions, markups decline.

Table 2 reports the resutls from the trend-cycle decomposition of changes in annual markups based on the trend-cycle estimation for annual elasticities provided in Section 3.3. The average trend growth in markups has been equal to 3.3% per year over 2001-2020. The impacts of negative aggregate demand shocks vary between -17.8% (for 2017-2018) and -8.0% (for 2008-2009).[23] In light of the trend-cycle decomposition of annual elasticities in Figure 6, the drastic reduction in elasticity (and a corresponding jump in markup) in the first year of the COVID-19 pandemic, 2020, is likely a short-run phenomenon driven by strong fiscal and monetary stimuli. People's concerns over getting COVID during multi-stop shopping may make them less price sensitive and thus raise stores' markups.

[23]Note that these cyclical changes are computed relative to the trend. That is, when calculating cyclical changes in annual markups, we subtract the estimated trend component from the raw changes in annual markups. Nevertheless, these raw changes in annual markups are also statistically significant right after negative aggregate demand shocks, which is not reported here but partly visible in Figure 8.

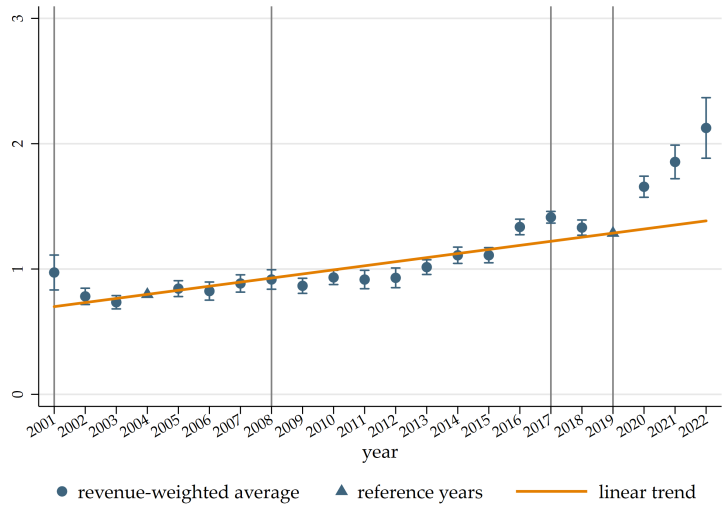


Figure 8: Trend and Cyclical Variation of Implied Markups.

Note: This figure shows the trend and cyclical variation of markups in the pooled IRI and NielsenIQ sample, where we shift IRI estimates to match the NielsenIQ average during the overlap period (same as Figure 7). Details of our estimation are discussed in the text. The four vertical gray lines indicate aggregate demand shocks. The spikes show the 95% confidence intervals of the revenue-weighted average annual demand elasticities relative to the linear trend. Standard errors are clustered at the market level.

Table 2: Trend and Cyclical Variation in Elasticity and Implied Markup.

	elasticity[a]	markup[b]
Linear trend (pre-COVID)		
average annual change, 2001-2019	-0.030***	0.033***
	(0.004)	(0.006)
Aggregate demand contractions (pre-COVID)		
from 2001 to 2002	0.419***	-0.223***
	(0.081)	(0.056)
from 2008 to 2009	0.137***	-0.083***
	(0.019)	(0.027)
from 2017 to 2018	0.083***	-0.115***
	(0.014)	(0.027)
Post-COVID annual changes		
from 2019 to 2020	-0.240***	0.338***
	(0.014)	(0.041)
from 2021 to 2022	-0.076***	0.239***
	(0.021)	(0.072)

Note: [a]This table reports estimates of the linear trend component as well as specific yearly changes in elasticities and markups during our sample period. The results in the table are based on the trend and cycle decompositions performed in Figures 6 and 8. Standard errors are clustered at the market level and listed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 Decomposition of markups dynamics

4.1 Cross-sectional variation in cyclical patterns

Results in the previous section show that elasticities and markups exhibit substantial secular and cyclical variation. In particular, it is evident that average markups react to aggregate demand shocks. At the same time, different markets follow heterogeneous patterns that may allow us to estimate the underlying causes of the time variation. For example, there may be heterogeneous shocks to local income, unemployment or housing prices.

There have been several mechanisms discussed in the literature that can drive price elasticities and markups. In particular, there is suggestive evidence that, with higher wealth, people become less sensitive to price changes (Stroebel and Vavra, 2019). Alternatively, markups may grow as a result of increased market concentration (Philippon, 2019). Elasticities and therefore markups may also vary along with income changes during business cycles (Aguiar et al., 2013).

The panel structure of our dataset gives us the opportunity to study the causes of the secular trend and cyclical dynamics in the markups. To gain a thorough understanding of possible underlying mechanisms, we next investigate cross-sectional variation in aggregate demand measures and indicators of market concentration.

4.2 Partial adjustment model for markups

Implied markups are a function of demand elasticities. It is well known that demand functions may gradually adjust to external factors due to, for example, habit formation. We therefore use the partial adjustment dynamic panel model of Balestra and Nerlove (1966) and Houthakker et al. (1974). An identifying assumption of this model is that the explanatory variables are weakly exogenous. Furthermore, this model distinguishes between short-run and long-run reactions of the dependent variable to changes in the explanatory factors.

We pool market-category-year-specific markups in the IRI and NielsenIQ samples and run a dynamic panel regression (5) below:

$$\hat{\mu}_{rjt} = \rho\hat{\mu}_{rj(t-1)} + \sum_{\ell=0}^{L} \left(\alpha_\ell R_{t-\ell} + M_{r(t-\ell)}\beta_\ell + C_{rj(t-\ell)}\gamma_\ell + D_t^{20} V_{r(t-\ell)}\theta_\ell\right) + c_j + m_r + D_t^{20} m_r + \epsilon_{rjt}. \quad (5)$$

We include several macroeconomic, demographic, and grocery market factors to explore drivers of demand elasticity and markup dynamics: R_t is the economy-wide real interest rate, M_{rt} are market-specific factors (real GDP per capita, unemployment rate, real housing prices, population, the number of employees per grocery store, the number of stores per capita), C_{rjt} are food-category-specific variable (the average number of UPC per category across stores

within each market). We also add category and market fixed effects, c_j and m_j, and the pandemic dummy variable $D_t^{20} = 1\{t \geq 2020\}$, and a subset of macro variables V_t (real GDP per capita and the unemployment rate).[24] The interaction terms $D_t^{20} V_{r(t-\ell)}$ account for the fact that the grocery sector has experienced a boom in spite of the economy-wide recession since the beginning of the pandemic as people shifted their consumption patterns in favor of food at home.

It is also important to mention that both average markups and the explanatory factors are persistent (see Figures 7 and Appendix Figures C.1 and C.2), which may cause inconsistent estimators due to the spurious correlation effect in a static model. Assuming that the trending variables have at most one unit root, we can restore consistency by including both lagged dependent and explanatory variables into the model (Sims et al., 1990). We set $L = 1$ so that the lagged dependent and explanatory variables are included.

As pointed out earlier, the partial adjustment model allows us to estimate both the short- and long-run effects of different variables. The short-run effects are equal to $(\alpha_0, \beta_0, \gamma_0)$ and capture contemporaneous effects of exogenous factors. The long-run effects are equal to $(\sum_{\ell=0}^{L} \alpha_\ell/(1-\rho), \sum_{\ell=0}^{L} \beta_\ell/(1-\rho), \sum_{\ell=0}^{L} \gamma_\ell/(1-\rho))$. They capture the predicted changes in the dependent variable in response to permanent shifts in the explanatory variables.

Another feature of this setup is that we can perform a historical decomposition in which we analyze the effect of variation in different explanatory variables on the dependent variable. For example, what was the effect over time of the real interest rate on markup dynamics? We will perform this exercise after discussing the explanatory variables.

4.3 Data on explanatory variables

Before proceeding, we briefly discuss the explanatory variables. We consider four types of changes in local grocery demand: (i) slow-moving demographic factors, (ii) changes in the competitive landscape of the local market, (iii) cyclical demand shocks driven by business cycle variation, and (iv) effect of monetary policy. All of these factors may affect local demand, and, as such, have an effect on grocery store price elasticities. This data is gathered from publicly available sources in addition to our scanner datasets. Appendix C lists data sources and describes the construction of all the explanatory variables. Here, we provide a brief discussion of their potential impacts on demand elasticities.

(i) Population: Local population growth may reflect consumer's outlook on future economic development in each market. A better economic outlook may give households more confidence in their future income growth and make them less price sensitive. Population growth may also boost the economy and thus increase income in the near future, again po-

[24] IRI and NielsenIQ have different product category names in letters, which are captured by our category fixed effects. Also, Table B.8 shows that our regression results remain almost unchanged when a linear-time trend, which may have affected all factors at the same time, is included in the econometric model.

tentially resulting in lower price sensitivity. We note that the population data show that, after the COVID-19 pandemic, some states, like California, saw a decline in population, while others, like Texas, saw an increase.

(ii.a) Average No. of UPCs per product category across stores: We calculate this measure at the market-category-year level based on the IRI and NielsenIQ scanner datasets. For each category and year, we take a weighted average of UPC numbers across stores within each market using store-category-specific revenue shares as weights. Intuitively, more UPCs in a product category mean more product varieties for consumers to choose from, which tends to make consumers more price sensitive. However, new products may be better tailored to consumers with heterogeneous tastes, which tends to make consumers less sensitive to price changes. In theory, these two effects may offset each other (Neiman and Vavra, 2023). We will use the data to analyze which effect dominates.

(ii.b) Grocery establishments per 1,000 residents: A higher number of establishments per capita means that it will be less costly for households to shop around for cheaper prices. Hence, they may become more price sensitive. However, more stores in a local grocery market may also mean less concentration and thus more competition (Philippon, 2019). If this is the case, then households may have less need to shop around for cheaper prices. We rely on the data to figure out which effect dominates or if they cancel out each other.

(ii.c) No. of employees per grocery store: Store sizes usually change along with the number of stores in a market. In particular, store sizes may increase with the reduction of store numbers, as shown by the County Business Patterns data (see Figure C.1). With more employees, larger stores may provide consumers with better services and more products. Consumers may then feel less need to shop around and thereby become less price sensitive.

(ii.d) Herfindahl-Hirschman index: HHI is a standard way to measure market concentration (e.g., Neiman and Vavra, 2023). We construct a retailer-HHI for each market during each year using NielsenIQ's consumer panel dataset.[25] In principle, a higher HHI means a higher market concentration so that consumers have more restricted retailer choices, resulting in lower price elasticities of demand.

(iii.a) Real GDP per capita: We use this as a measure of real income. Increases in household income may result in lower price sensitivity.

(iii.b) Unemployment rate: The unemployment rate is also related to real income but may also affect households' available time. Households experiencing unemployment are more likely to shop around for lower prices (Aguiar et al., 2013).

(iii.c) Cumulative percentage changes in real housing prices: For homeowners, rising housing prices can mean higher wealth and therefore make them less price sensitive (Stroebel

[25]Following Neiman and Vavra (2023), we use household-year-specific weights, provided by NielsenIQ, to calculate the retailer-specific market shares of each year, together with household-retailer-year-specific expenditures. We do not calculate HHI's for the IRI sample as it does not have the required data.

and Vavra, 2019). However, Stroebel and Vavra (2019) also found that, for renters, higher housing prices can mean higher rent and therefore make them more sensitive to grocery prices. Which effect dominates is an empirical question.

(iv) Real interest rate: We use the one-year real interest rate from FRED.[26] The interest rate is the key determinant of current consumption spending decisions in modern macroeconomic models (Woodford, 2003). Higher interest rates incentivize more savings at the expense of current consumption. Moreover, in the short run, the Fed can adjust the real interest rate using standard monetary policy tools.

In Figures C.1 and C.2 in the appendix, we show the variations of all these factors. Importantly, with the exception of the real interest rate variable, all other factors have both market-specific and time variation. In the partial adjustment model estimation we exploit this statistical power to identify the impacts of these factors on demand elasticities and therefore markups across markets.[27]

4.4 Estimation results

We use WLS to estimate the model in equation (5). For each year, we use market-category-specific revenue shares as weights within the IRI or NielsenIQ sample. For the overlapping period of 2006-2012, we adjust IRI and NielsenIQ sample weights using their observation shares within each year.

As discussed in Section 2.2, it is important to use revenue share weights to capture the average effects of the factors on the markups of a typical consumer basket. Indeed, food groceries are very heterogeneous and the revenue shares reflect relative spending on different categories.

Although our dynamic panel is relatively long ($T = 22$), the regression coefficient estimates are still subject to $O(1/T)$ bias, as pointed out by Nickell (1981).[28] However, recent work by Klosin (2024) shows that such a dynamic bias is typically much smaller than the bias from alternative estimation methods such as the static panel model. We cluster standard errors at the market level to allow for arbitrary correlations among regression errors between product categories and time within each market.

Table 3 shows the results. We find that macroeconomic, demographic, and grocery market factors can explain a larger share of the trend and cyclical variation of implied markups.[29] We report two sets of results. In columns (1) and (3) we report short-term effects, while

[26] See: https://fred.stlouisfed.org/series/REAINTRATREARAT1YE.

[27] Owyang et al. (2005) and Baumeister et al. (2022) find substantial differences in the timing of business cycles across the U.S.

[28] A typical standard error of an estimator in a time series application is $O(1/\sqrt{T})$. So, the second order bias in our application is comparable to a standard error with $T = 22^2 = 484$ observations, which is common in modern empirical practice.

[29] Table B.7 presents the corresponding results for demand elasticities.

columns (2) and (4) report the corresponding long-run effects of the different factors. The adjusted R^2 is a whole-regression R^2 that takes into account degrees of freedom or the number of all variables (including fixed effects). The within R^2 measures the explanation power of our factor variables (excluding fixed effects). Of the 48% within R^2, 25% and 27% (columns (1) and (3)) are due to the explanatory variables apart from the lagged dependent variable. This highlights the fact that it is important to model markups in a dynamic model, while also indicating that the set of included variables explain a large share of the variation.

We find that higher (log) population increases markups, in particular in the long-run. The estimate is a little noisy, possibly because it tends to be slow-moving and may be correlated with other variables, e.g. real GDP.

Grocery-market specific factors also affect markups. The number of varieties and retailer concentration in a specific market are statistically significant. More varieties in a food category lead to lower markups. This implies that the substitution effect away from a specific product and the resulting increase in competition dominates. The concentration of stores in a market, for example, if a particular chain has a larger market share, leads to higher markups. The number of establishments and the size of the store as measured by the No. of employees are not statistically significant after accounting for other factors.

Table 3: The Factor Regression Results for Implied Markups.

explanatory variables[a]	dependent variable: *implied markup*[b] 2001-2022		2004-2022	
	(1)	(2)	(3)	(4)
	short-term	long-run[c]	short-term	long-run
lagged implied markup	0.57***		0.57***	
	(0.02)		(0.02)	
real interest rate	-5.92***	-9.82***	-7.74***	-9.97***
	(1.04)	(2.01)	(0.88)	(2.22)
real GDP per capita (pre-COVID)	1.21**	2.45***	1.65***	2.20***
	(0.44)	(0.51)	(0.42)	(0.60)
unemployment rate (pre-COVID)	-0.08	-1.65	0.84	-3.90***
	(0.65)	(1.14)	(0.65)	(1.21)
cum. perct change in real housing price	0.32***	0.27**	0.31***	0.02
	(0.07)	(0.13)	(0.06)	(0.16)
population	0.02	0.88	0.19	1.13*
	(0.22)	(0.63)	(0.25)	(0.60)
No. of UPCs per category	-0.17***	-0.25***	-0.18***	-0.33***
	(0.04)	(0.07)	(0.03)	(0.08)
grocery establishments per 1k residents	-0.18	-0.07		
	(0.29)	(0.49)		
No. of employees per grocery store	-0.03	-0.13		
	(0.03)	(0.10)		
retailer HHI			0.24**	0.23**
			(0.10)	(0.11)
post-COVID controls[d]				
interaction with real GDP per capita	-5.73***		-6.41***	
	(1.76)		(1.78)	
interaction with unemployment rate	-10.21***		-12.14***	
	(3.08)		(3.14)	
interaction with market dummies	YES		YES	
one lag of all variables (no dummies)	YES		YES	
fixed effects of markets and goods	YES		YES	
$adj.R^2$	0.66		0.65	
within R^2	0.48		0.48	
N	23,322		22,527	

Note: [a]Explanatory variables are logged except for real interest rate, unemployment rate, and cumulative percent change in real housing price. See details in the main text and Appendix C. [b]We obtain implied markups from equation (3); these markups are market-category-year-specific and in levels. In the calculation, we shifted IRI elasticity estimates down by a constant such that the IRI and NielsenIQ samples have the same markup mean for the overlapping period of 2006-2012. [c]Given the distributed lag model, we impute the long-run effects for each factor and estimate standard errors using the Delta method. [d]Pandemic controls have two parts: one captures heterogeneous level shifts across markets and the other captures interactions with GDP and unemployment. All the models above are estimated by OLS using market-category-specific revenue shares as weights within each year. Standard errors, clustered at the market level, are listed in the parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the group of macro effects, real per capita GDP and housing prices have the largest effects. Higher income leads to higher markups and higher housing prices do the same, indicating that the wealth effect of homeowners may dominate the effect of potentially higher rents. The effect of unemployment is estimated with lower precision, most likely due to the high correlation with other variables. The one exception is the long-run effect in column (4), which indicates that higher unemployment results in lower markups. Meanwhile, housing prices become insignificant in this specification.[30] In the next section, when we perform a historical decomposition of the variation in markups we group variables together and compute the overall effect of these variables. One reason for this approach is the within-group correlation across variables.

The real interest rate has a negative effect on markups. This effect is consistent with the standard consumption and savings decision where a higher real rate results in lower demand today. Our results indicate that markups decline when the real interest rate increases. We note that this effect is present in the short-run but is even stronger in the long-run. If we assume that prices are not sticky over longer time horizons, our result suggests a direct effect of monetary policy on grocery store prices. This result is relevant for the literature on the transmission mechanism of monetary policy.

The actual economic impact of any specific variable on markups depends both on the estimated coefficient but also on the variation of the explanatory variable. In the next section we therefore perform a historical decomposition.

Finally, we find evidence of a structural break in the effect of unemployment and per capita GDP when the COVID-19 pandemic hits in 2020. Typically GDP and unemployment are negatively correlated and the effect of GDP is stronger. However, during COVID, the unemployment becomes much more dominant, perhaps because unemployment during this period may be proxying for lockdowns. The conditional correlation appears to be even larger since the sign on per capita GDP has now switched.

As a robustness check, we also report the markup estimation results when we include a linear time trend in the regression (see Table B.8). The coefficient for the linear time trend is not statistically significant. This justifies our preferred specification in equation (5). Moreover, we find almost the same results as reported here.

4.5 Factor decomposition

It is interesting to study the contribution of different groups of factors on the overall and market-specific dynamics. We therefore perform the following historical decomposition.

[30] We note that Stroebel and Vavra (2019) find a related wealth-effect mechanism. Their argument is based on the assumption that food retail costs (e.g., store rents) do not react to changes in housing prices, thus the positive effect of higher housing prices on grocery prices from homeowners leads to higher retail markups.

For each market, we calculate the difference in predicted values under scenarios in which a given factor remains constant at the initial level and the one where the factor follows the historical trajectory. The purpose of this exercise is to measure the effect of different sets of explanatory variables—demographic, market-specific, macroeconomic, and real interest rate—on the average markups in various markets while accounting for the transition dynamics.

Figure 9 Panel A reports the nationwide average markups, across all of the markets in our sample, together with the markups predicted by the variation in the explanatory variables. We use results from the specification in Table 3 column (1) to calculate markups; we revenue-weight across categories within a market and then across markets.[31]

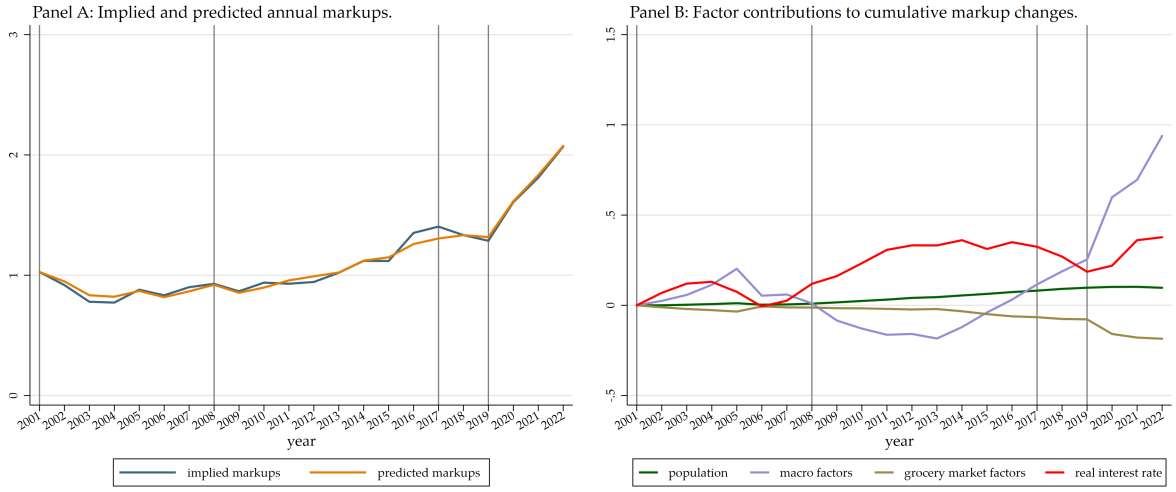


Figure 9: Revenue-weighted Annual Markups.

Note: We first obtain predicted market-category-year-specific markups based on the estimated regression model (5). We then calculate revenue-weighted annual markups within the IRI and NielsenIQ samples, respectively. For the overlapping period of 2006-2012, we average annual markups using the IRI and NielsenIQ sample size shares as weights. We follow the same procedures for the annual markups implied from elasticity estimates. To calculate factor contributions, we first set all factors at their initial values to obtain predicted annual markups under self-dynamics, which are used as the baseline predicted annual markups. We then calculate the cumulative contributions of each group of factors by taking the cumulative differences between the factor-predicted annual markups and the baseline predicted annual markups. We group our eight factors into four groups: (1) real GDP per capita, unemployment rate, and cumulative changes in real housing prices combined as macro factors together with pandemic effects; (2) the number of grocery stores per capita, the number of employees per grocery store, and the number of UPCs per product category combined as grocery market factors; (3) population; and (4) real interest rate.

Figure 9 Panel B shows the contribution of population, macro factors, grocery market factors, and the real interest rate. We note a slow upward trend in markups due to population and a slow downward trend due to grocery market factors. This implies that the changes in

[31]Figure C.3 in the appendix shows the alternative specification of the grocery market factors that includes HHI over a shorter sample of 2004-2022.

the competitive landscape put downward pressure on markups over the sample period. This result suggests that the grocery retail sector became more competitive.

Higher frequency cyclical changes are mainly driven by macro factors and the real interest rate as well as transitional dynamics. For example, in 2008, there is a negative effect of macro factors on markups, and in 2017 there is downward pressure on markups due to higher real interest rates.

Overall, we see evidence of downward pressure of macro factors on markups from 2009 to 2014, a period of somewhat sluggish growth after the Great Recession. At the same time we see the supportive effect of accommodative monetary policy during the period. We also see a large increase in markups from 2019 to 2020, which at least partly was driven by COVID specific developments captured by the macro factors.

Figure 9 Panel A shows a very high fit. Markups are averages across markets and categories. As we are about to discuss, we see similarly high levels of fit for individual markets, implying that the unexplained variation is primarily due to differences across categories.

In our panel data set all variables, except the real interest rate, vary at least at the market level. That is, real GDP growth is different in LA and Chicago. This setup therefore allows us to explore the ability of our model to explain regional variation in markups. We therefore consider the largest markets from the four macro-regions of the country – NYC (East), Chicago (Midwest), Dallas (South) and LA (West). Figure 10 shows overall actual and fitted markups as well as factor-specific historical decompositions.

We find that population dynamics generally have a very small impact on markups, except for Dallas. The effect of grocery market factors is similar across markets. The largest driver of differences across markets are the macro factors – per capita GDP growth, unemployment, housing prices, and the effects of COVID. The largest post-Great-Recession impact was in LA, which also saw the highest recent increase in markups due to macro factors. Chicago, which experienced only small changes in macro factors saw the lowest overall increase in markups during the sample period.

We note that, in the figure, the real interest rate had a slightly different reported effect across markets. This is because of compositional effects in the unbalanced panel. Therefore, the average effects of common factors are not exactly aligned across markets.

The results imply that our model does a good job explaining variation in average markups over time across markets. We also note that the real interest rate plays a very important role in explainig variation in markups, a point which we discuss in more detail below.

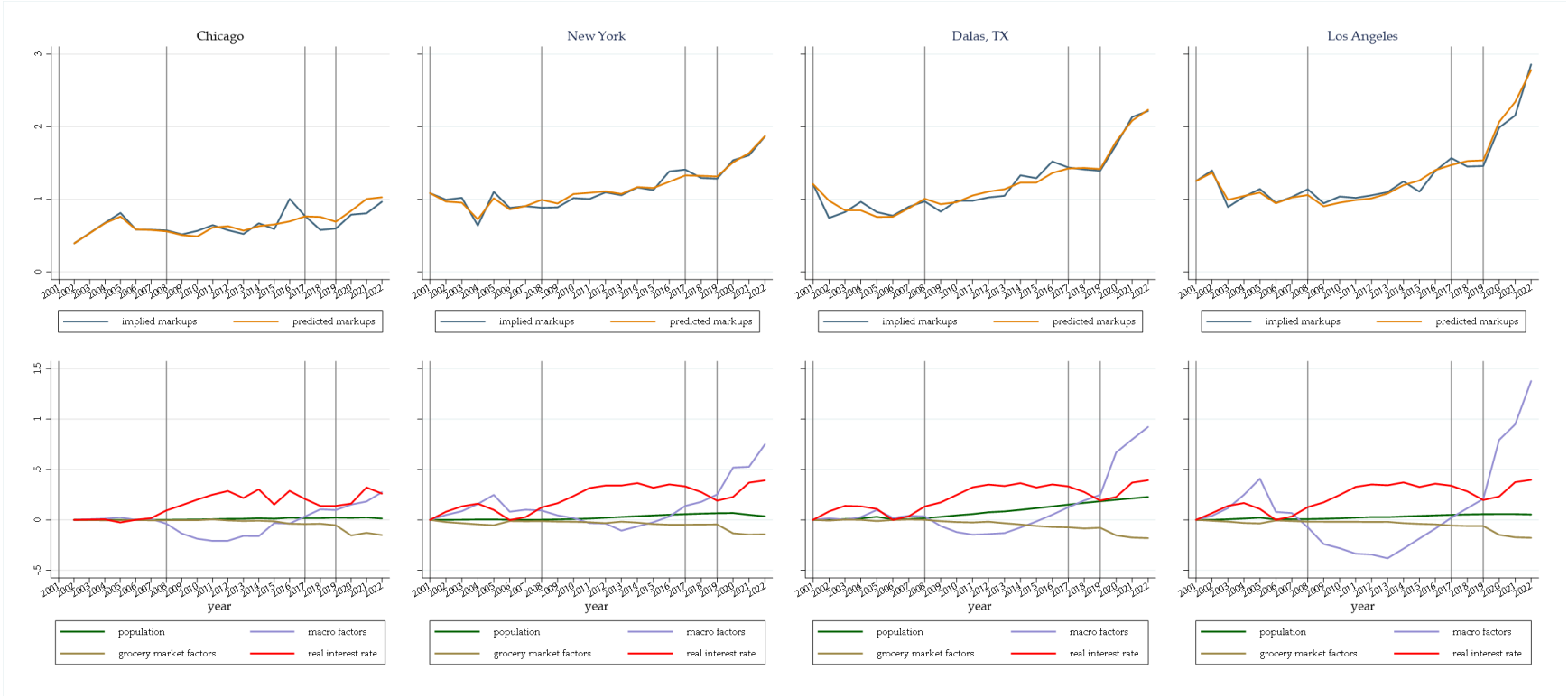


Figure 10: Revenue-weighted Annual Markups of the Four Largest Regional Markets.

5 Discussion

5.1 Implication for long run movements in food costs and inflation

According to our estimates, the average markup growth from 2006 to 2019 is about 54.4%. This number is comparable with other studies, e.g., De Loecker et al. (2020), Brand (2021), and Döpper et al. (2022). The average markup growth from 2019 to 2022 was even higher, equal to 60.9%. This is in line with the accelerated inflation during the later part of the sample.[32] Overall, the average markup increased by 101.5% from 2001 to 2022.

We next explore the implications of the large increase in average markups on food inflation. We consider a simple baseline scenario where we assume that nominal average marginal costs c_* stayed constant throughout the sample period. In this case, food prices would have increased from $(1+\mu_{2006})c_*$ to $(1+\mu_{2022})c_*$, i.e. by 51.4%. We can now compare this benchmark increase in food prices – where everything is attributed to markup growth – to actual increases in food prices. The light blue area on Figure 11 reports the nation-wide food at home CPI index.[33] Interestingly, the CPI subindex for food at home (food groceries) has increased by 66.4% over the same period, which corresponds to a below 1% difference per year. The most likely explanation for this pattern is that the nominal marginal cost of food has remained approximately unchanged, i.e., the real marginal cost has declined during this period. Indeed, Döpper et al. (2022) note that this may have been why larger markups did not reach consumers in the form of higher real food prices.

In order to explore this hypothesis further, we calculate the implied marginal costs obtained using our markup estimates. Each year, we deflate the food at home CPI using our average markup estimate and calculate the implied nominal marginal cost for food. The dark blue area in Figure 11 shows this series. We note that it indeed remained almost constant over the sample period.

Over our sample period, the real cost of food at grocery stores did not change – the increase in food at home costs (66.4%) and the overall CPI increased by close to the same amount (68.4%). However, our results suggest that real marginal costs of food declined. It is possible that this lower cost was the result of large fixed-cost investments and efficiency increases in the food industry (as suggested, for example in Watson and Winfree (2022)). Such investments may have resulted in large improvements in logistics and food waste management. Meanwhile, labor productivity in the food retail industry increased by 41.9% from 2001 to 2022.[34] This

[32] In related work, Gafarov et al., 2024, find that markup estimates can significantly explain the variation in fresh produce grocery prices.

[33] The overall CPI index refers to the "Consumer Price Index for all urban consumers: all items in U.S., city average" and the CPI subindex for food is the "Consumer Price Index for all urban consumers: food at home in U.S., city average". We obtain both data from FRED.

[34] The labor productivity among US food and beverage stores has increased from 77.7 (an index) in 2001 to 110.2 in 2022, according to the FRED data (https://fred.stlouisfed.org/series/IPUHN445L000000000).

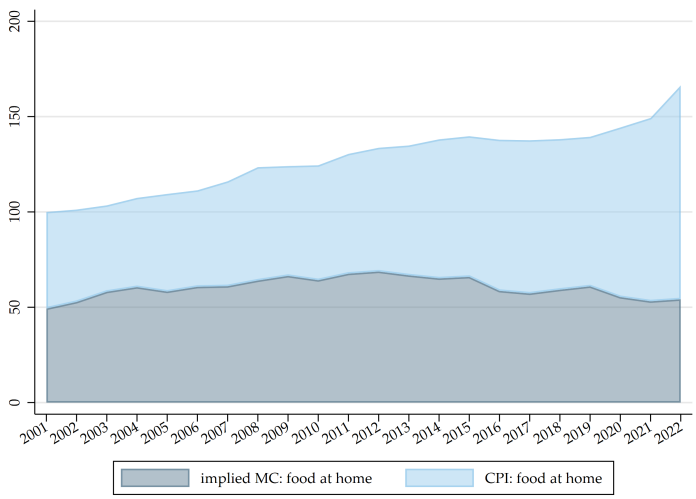


Figure 11: Food-at-home CPI and Implied Marginal Cost.

Note: We obtain food-at-home CPI from FRED. Given our annual implied markups, we calculate the implied nominal marginal costs following equation (3).

higher productivity suggests that grocery stores have been providing more comprehensive and possibly higher-quality service to consumers. While not necessarily affecting marginal costs, this may have increased average costs, e.g. advertising and marketing. Thus, higher service quality may have contributed to a higher willingness to pay by customers and, in turn, covered the increased average costs faced by grocery stores. We leave the exploration of these patterns and hypotheses for future research.

5.2 Evidence of procyclical markups

One of our main results is identifying procyclical movement in markups. This finding is related to a growing empirical literature on procyclical markups as summarized in Nekarda and Ramey (2020). Complementing earlier work, our paper provides micro-level evidence on the procyclicality of optimal markups (the markups in the absence of sticky prices), which our results suggest is largely driven by income effects.

These patterns are in contrast to earlier theoretical models of countercyclical markups (Rotemberg and Saloner, 1986; Rotemberg and Woodford, 1999). The comovement in these papers underlies the transmission mechanism for the effects of demand shocks in sticky-price New Keynesian models. The idea is that lower interest rates are associated with higher demand and therefore lower markups. However, this theoretical pattern is at odds with our findings and the results in Nekarda and Ramey (2020). Instead, markups increase in response to positive demand shocks, a pattern they refer to as conditional procyclicality.

Based on the results from our structural dynamic panel estimation, we find effects of demand shocks and the interest rate. In our estimation approach we can distinguish between

these two effects using cross-sectional variation in demand shocks.

As shown in Table 2, the monetary tightening in 2017 coincided with an average decrease of 11.5% in markups, while the government spending stimulus in 2020 was associated with an average increase of 33.8% in markups. These effects are unconditional responses. The conditional responses in the table reflect different effects potentially offsetting one another. For example, the negative aggregate demand shock in 2008 resulted in only a small change in markups. Our estimates imply that loose monetary policy counteracted low aggregate demand. To be specific, if we combine population, grocery market and macro factors from 2007 to 2013, the cumulative effect on markups would have been -23.6%. Adding in the interest factor instead resulted in a cumulative change of 10.4% (see Figure 9). Our results therefore imply an important role of monetary policy in stabilizing prices through their effect on markups.

6 Conclusion

In this paper, we make two main contributions to the literature on markup estimation and the cyclicality of markups and competition.

First, we estimate revenue-weighted demand elasticities in a panel-IV regressions model that implements the widely-used Hausman IV strategy. Our findings are non-parametric in the sense that we do not assume any single particular parametric model of consumer demand. Our particular implementation only assumes common wholesale suppliers for each specific pair of geographically close markets. Unlike some others, we work directly with weekly sales data which results in strong IVs.

Second, we provide direct micro-level evidence of a substantial downward trend in demand elasticities and an associated upward trend in markups in the US from 2001 to 2022. Moreover, our markup measure exhibits economically large and statistically significant drops at times of contractions in aggregate demand, implying procyclical variation in markups. The patterns in food markups that we find are representative of the whole US economy since we cover markets in the Western, Mid-Western, Eastern, and Southern US.

We use a dynamic adjustment model and a calculate causal effects of the competitive landscape and consumer factors on markups. We then conduct a structural decomposition to attribute changes in markups across regions and over time to various factors. Our structural attribution allows us to answer the question we asked at the beginning: Does policy intervention supporting aggregate demand increase competition and welfare? We find that increases in aggregate demand factors and accommodative monetary policy both have strong positive effects on markups. The implication is that competition is lower in times of aggregate demand expansion. We also find that proxies for the competitive landscape affect markups—when competition increases, markups decline. Although there is some suggestive evidence

that competition may, at times, increase when aggregate demand increases, this effect is not strong enough to counteract the positive effect of aggregate demand and monetary policy on markups.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-based versus design-based uncertainty in regression analysis," *Econometrica*, 88, 265–296.

Afrouzi, H. and L. Caloi (2024): "Endogenous firm competition and the cyclicality of markups," *Review of Economics and Statistics*, 106, 1638–1654.

Aguiar, M., E. Hurst, and L. Karabarbounis (2013): "Time use during the great recession," *American Economic Review*, 103, 1664–1696.

Anderson, E., S. Rebelo, and A. Wong (2023): "Markups across space and time," *Unpublished Manuscript*.

Backus, M., C. Conlon, and M. Sinkinson (2021): "Common ownership and competition in the ready-to-eat cereal industry," Tech. rep., National Bureau of Economic Research.

Balestra, P. and M. Nerlove (1966): "Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas," *Econometrica: Journal of the econometric society*, 585–612.

Baumeister, C., D. Leiva-León, and E. Sims (2022): "Tracking weekly state-level economic conditions," *Review of Economics and Statistics*, 1–45.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890.

Bogin, A., W. Doerner, and W. Larson (2019): "Local house price dynamics: New indices and stylized facts," *Real Estate Economics*, 47, 365–398.

Bonhomme, S. and E. Manresa (2015): "Grouped patterns of heterogeneity in panel data," *Econometrica*, 83, 1147–1184.

Brand, J. (2021): "Differences in differentiation: Rising variety and markups in retail food stores," *Available at SSRN 3712513*.

Broer, T., N.-J. Harbo Hansen, P. Krusell, and E. Öberg (2020): "The New Keynesian transmission mechanism: A heterogeneous-agent perspective," *The Review of Economic Studies*, 87, 77–101.

Bronnenberg, B. J., M. W. Kruger, and C. F. Mela (2008): "Database paper—The IRI marketing data set," *Marketing Science*, 27, 745–748.

Bugni, F. A., I. A. Canay, and S. McBride (2023): "Decomposition and interpretation of treatment effects in settings with delayed outcomes," *arXiv preprint arXiv:2302.11505*.

Chetverikov, D. and E. Manresa (2022): "Spectral and post-spectral estimators for grouped panel data models," *arXiv preprint arXiv:2212.13324*.

Chevalier, J. A. and D. S. Scharfstein (1995): "Liquidity constraints and the cyclical behavior of markups," *The American Economic Review*, 85, 390–396.

Cournot, A.-A. (1838): *Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot*, chez L. Hachette.

Cox, C. and K. Harris-Lagoudakis (2022): "SNAP eligible products and behavioral demand," *Available at SSRN 4146549*.

D'Acunto, F., U. Malmendier, J. Ospina, and M. Weber (2021): "Exposure to grocery prices and inflation expectations," *Journal of Political Economy*, 129, 1615–1639.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 135, 561–644.

De Loecker, J. and P. Scott (2022): "Markup estimation using production and demand data. an application to the us brewing industry," *Unpublished paper*.

De Loecker, J. and P. T. Scott (2016): "Estimating market power: Evidence from the US brewing industry," Tech. rep., National Bureau of Economic Research.

DellaVigna, S. and M. Gentzkow (2019): "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 134, 2011–2084.

Döpper, H., A. MacKay, N. Miller, and J. Stiebale (2022): "Rising markups and the role of consumer preferences," *Harvard Business School Strategy Unit Working Paper*.

Gaarder, I., L. H. de Frahan, M. Mogstad, A. Torgovitsky, and O. Volpe (2024): "Linear Supply and Demand in Heterogeneous Markets," .

Gafarov, B., T. Gong, and J. Hilscher (2024): "Do organic products have higher markups?" *UC Davis working paper*.

Gallego, G., H. Topaloglu, et al. (2019): *Revenue management and pricing analytics*, vol. 209, Springer.

Gandhi, A., S. Navarro, and D. A. Rivers (2020): "On the identification of gross output production functions," *Journal of Political Economy*, 128, 2973–3016.

Harris-Lagoudakis, K. (2023): "The effect of online shopping channels on brand choice, product exploration and price elasticities," *International Journal of Industrial Organization*, 87, 102918.

Hausman, J. A. (1996): "Valuation of new goods under perfect and imperfect competition," in *The economics of new goods*, University of Chicago Press, 207–248.

Houthakker, H. S., P. K. Verleger, and D. P. Sheehan (1974): "Dynamic demand analyses for gasoline and residential electricity," *American Journal of Agricultural Economics*, 56, 412–418.

Imbens, G. W. and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica: Journal of the Econometric Society*, 467–475.

Keynes, J. M. (1936): "The general theory of interest, employment and money," .

Kilian, L. and H. Lütkepohl (2017): *Structural vector autoregressive analysis*, Cambridge University Press.

Klosin, S. (2024): "Dynamic biases of static panel data estimators," *arXiv preprint arXiv:2410.16112*.

Lerner, A. (1934): "The concept of monopoly and the measurement of monopoly power," *The Review of Economic Studies*, 1, 157–175.

Levin, L., M. S. Lewis, and F. A. Wolak (2017): "High frequency evidence on the demand for gasoline," *American Economic Journal: Economic Policy*, 9, 314–347.

MacKay, A. and N. Miller (2023): "Estimating models of supply and demand: Instruments and covariance restrictions," *Harvard Business School Strategy Unit Working Paper*.

Munro, E., S. Wager, and K. Xu (2023): "Treatment effects in market equilibrium," *arXiv preprint arXiv:2109.11647*.

Neiman, B. and J. Vavra (2023): "The rise of niche consumption," *American Economic Journal: Macroeconomics*, 15, 224–264.

Nekarda, C. J. and V. A. Ramey (2020): "The cyclical behavior of the price-cost markup," *Journal of Money, Credit and Banking*, 52, 319–353.

Nevo, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Nickell, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica: Journal of the econometric society*, 1417–1426.

Owyang, M. T., J. Piger, and H. J. Wall (2005): "Business cycle phases in US states," *Review of Economics and Statistics*, 87, 604–616.

Phelps, E. S. and S. G. Winter (1970): "Optimal price policy under atomistic competition," *Microeconomic foundations of employment and inflation theory*, 309–337.

Philippon, T. (2019): *The great reversal*, Harvard University Press.

Richards, T. J. and S. F. Hamilton (2015): "Variety pass-through: An examination of the ready-to-eat breakfast cereal market," *Review of Economics and Statistics*, 97, 166–180.

Rotemberg, J. J. and G. Saloner (1986): "A supergame-theoretic model of price wars during booms," *The American economic review*, 76, 390–407.

Rotemberg, J. J. and M. Woodford (1992): "Oligopolistic pricing and the effects of aggregate demand on economic activity," *Journal of political Economy*, 100, 1153–1207.

——— (1999): "The cyclical behavior of prices and costs," *Handbook of macroeconomics*, 1, 1051–1135.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in linear time series models with some unit roots," *Econometrica: Journal of the Econometric Society*, 113–144.

Staiger, D. and J. H. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

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

Watson, P. and J. Winfree (2022): "Should we use antitrust policies on big agriculture?" *Applied Economic Perspectives and Policy*, 44, 1313–1326.

White, H. (2014): *Asymptotic theory for econometricians*, Academic press.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.

Working, E. J. (1927): "What do statistical âdemand curvesâ show?" *The Quarterly Journal of Economics*, 41, 212–235.

Appendix

A Details on IRI and NielsenIQ datasets

For comparability, we do not take the lower product module in the NielsenIQ sample as the product category, since it has far fewer UPCs than a product category in the IRI sample does. Instead, we take food groups as the product category for the NielsenIQ sample. Table B.5 shows that these two samples now have a comparable number of UPCs per food category in magnitude. We also drop four food categories or groups in the NielsenIQ sample: ICE, YEAST, GUM, and PET FOOD. For the first two, our paired markets have a limited number of common products at the UPC level, which would make our IV strategy infeasible if we included these two food categories in our estimation sample. For the last two food categories, we drop them as we focus on edible human food in this paper.

Regarding UPCs, we take them as given in the IRI and NielsenIQ retail scanner data sets. Both data sets have collapsed private-label UPCs across retailers to some extent. The assumption is that the core attributes are identical for these UPCs, although there may be some custom characteristics that differ depending on how the different products are marketed. Using these collapsed UPCs and retailer codes, we can construct retailer-specific private-label UPCs within each food category. However, we find that it will not only weaken the strength of our price IVs but also lower our elasticity estimates. This suggests that refining private-label UPCs will not help us recover reasonable price elasticities of demand. Hence, we stick to the collapsed private-label UPCs in our main results. Results based on retailer-specific private-label UPCs are available upon request.

When it comes to the food stores of paired markets listed in Table B.1, we focus on panel stores that have reported product transactions across all 12 years in the IRI retail scanner data set and across all 17 years in the NielsenIQ retail scanner data set. Both data sets have store entries and exits over time. In particular, NielsenIQ's coverage rate for all US food stores decreased from 53% in 2011 to 26% in 2017. As a result, in our paired markets, the total number of food stores covered by NielsenIQ decreased by 24% from 2011 to 2017. We study time variation of annual markups among food stores that are covered in the data across all years. In this way, our results will not be mechanically driven by any changes in the store coverage of the data.

B Additional tables and figures

Table B.1 presents the specific pairs of the 25 markets shown on the US mainland map (Figure 1). Table B.2 reports the number of observations and stores across markets as well

as population for the IRI and NielsenIQ samples. In total, each year, the IRI data include close to 48 million observations from 429 stores, while the NielsenIQ data contain over 2,900 million observations from more than 1,700 stores. There are notable size differences across Markets. For example, the market with the most stores and observations, LOS ANGELES, is about 30 times larger than Syracuse, one of the smallest markets.

At the store level, Table B.3 shows that almost all panel stores report transaction records for most weeks. The average reported weeks per year-store are 51 and 52 for the IRI and NielsenIQ samples, respectively. Moreover, stores sell products for all the food categories of interest in this paper. Hence, we have sufficient observations nearly for every store. The average obervations per year are 113 and 602 thousands for the IRI and NielsenIQ samples, respectively, at the store level. As expected, stores sell fewer UPC products per category than markets. In the regression, we aggregate UPC revenues across stores and use UPC-specific revenue shares to estimate the revenue-weighted price elasticity of demand for each food category within each market and year.

Table B.1: The 12 Pairs of 25 Markets.

region	market A	market(s) B
EAST	BUFFALO/ROCHESTER[a]	SYRACUSE
	BOSTON	HARTFORD
	NEW YORK	PHILADELPHIA
SOUTH	CHARLOTTE	RALEIGH/DURHAM
	ATLANTA	SOUTH CAROLINA
	DALLAS, TX	HOUSTON
WEST	LOS ANGELES	SAN DIEGO
	SACRAMENTO	SAN FRANCISCO
	PORTLAND, OR	SEATTLE/TACOMA
MIDWEST	KANSAS CITY	DES MOINES + OMAHA[b]
	CHICAGO	PEORIA/SPRINGFLD.
	DETROIT	TOLEDO

Note: We take markets defined by IRI as given. Each IRI market consists of one or more adjacent counties. We use county-specific FIPS codes identify markets to which NielsenIQ stores belong. [a]The "/" means that the two areas belong to a single market. [b]The "+" means that the two areas belong to two markets but are combined as one by us when implementing our identification strategy. Specifically, when there is no single neighboring market with a sufficient number of stores, we pair a large market of interest with two smaller neighboring ones and utilize their average product prices to instrument the product prices of the large market of interest. However, when the two small markets are of interest, we treat them separately and use the product prices of the large market to instrument their product prices. Hence, we have 25 markets in total for demand elasticity estimations.

In the main text above, we focus on the revenue-weighted IV estimates of demand elasticities. Here, we report the summary stats for the equally-weighted IV estimates. Table B.6 shows that the equally-weighted elasticity estimates are sizably lower not only in mean but also in percentiles, such as the 10 and 90 percentiles of the distribution, compared to the revenue-weighted estimates as summarized in Table 1. However, their standard errors are slightly smaller and the first-stage F stats are mildly larger.

Table B.2: Summary Statistics of Markets.

market name	population in 2010 (million)	IRI (2001-2012) obs per year (million)	IRI (2001-2012) panel stores (count)	NielsenIQ (2006-2022) obs per year (million)	NielsenIQ (2006-2022) panel stores (count)
NEW YORK	19.5	4.9	41	148.9	262
LOS ANGELES	17.1	5.6	55	239.8	414
CHICAGO	9.0	1.9	16	123.6	179
PHILADELPHIA	6.5	3.0	24	57.8	109
DALLAS, TX	6.2	1.5	15	98.8	163
SAN FRANCISCO	6.1	2.4	23	100.6	174
HOUSTON	5.9	2.0	20	80.0	122
BOSTON	5.5	2.6	19	86.0	138
SOUTH CAROLINA	5.1	1.6	20	68.5	139
ATLANTA	4.9	2.2	17	82.9	116
DETROIT	4.8	0.9	7	57.9	79
SEATTLE/TACOMA	3.4	2.4	24	103.3	169
RALEIGH/DURHAM	3.3	1.8	19	119.1	236
HARTFORD	3.2	1.1	8	16.7	31
PORTLAND,OR	3.2	1.6	15	82.8	137
SAN DIEGO	3.1	1.8	16	43.8	72
SACRAMENTO	2.8	1.4	12	32.1	56
CHARLOTTE	2.6	1.3	13	88.4	176
BUFFALO/ROCHESTER	2.5	1.7	15	0.9	8
TOLEDO	2.0	0.9	7	21.3	29
PEORIA/SPRINGFLD.	2.0	0.6	5	17.8	27
KANSAS CITY	2.0	2.3	17	15.0	18
SYRACUSE	1.2	0.8	8	4.4	10
OMAHA	1.1	0.8	6	21.2	28
DES MOINES	0.7	0.9	7	15.6	22
total	123.8	47.8	429	1727.1	2914

Note: We report the store-UPC-week observations only for the food categories listed on Table B.4.

Table B.3: Summary Statistics of Stores.

	IRI stores			NielsenIQ stores		
	mean	p10	p90	mean	p10	p90
obs per year (thousand)	113.1	81.5	145.3	602.1	429.6	793.1
No. of weeks per year	50.9	48.8	52.1[a]	52.0	51.5	52.2[b]
No. of categories per year	16.0	16.0	16.0	58.0	58.0	58.0
No. of UPCs per year-category	222.3	171.8	272.2	367.1	269.4	471.8

Note: [a]A IRI week starts on Monday and ends on Sunday. [b]A NielsenIQ week starts on Sunday and ends on Saturday. Hence, both IRI and NielsenIQ stores have sales records of 53 weeks in some years.

Table B.4: The List of Food Categories.

IRI food (16)	NielsenIQ non-dry food (22)[a]	NielsenIQ dry food (36)[b]
carbonated beverages	BAKED GOODS-FROZEN	BABY FOOD
coffee	BREAKFAST FOODS-FROZEN	BAKING MIXES
cold cereal	BUTTER AND MARGARINE	BAKING SUPPLIES
frozen dinners&entrees	CHEESE	BREAD AND BAKED GOODS
frozen pizza	COT CHEESE, SOUR CREAM, TOPPINGS	BREAKFAST FOOD
hotdog	DESSERTS/FRUITS/TOPPINGS-FROZEN	CANDY
margarine butter	DOUGH PRODUCTS	**CARBONATED BEVERAGES**
mayonnaise	DRESSINGS/SALADS/PREP FOODS-DELI	CEREAL
milk	EGGS	**COFFEE**
mustard&ketchup	FRESH MEAT	CONDIMENTS, GRAVIES, AND SAUCES
peanut butter	FRESH PRODUCE	COOKIES
salty snacks	ICE CREAM, NOVELTIES	CRACKERS
soup	JUICES, DRINKS-FROZEN	DESSERTS, GELATINS, SYRUP
spaghetti&italian sauce	**MILK**	FLOUR
sugar substitute	PACKAGED MEATS-DELI	FRUIT - CANNED
yogurt	PIZZA/SNACKS/HORS DOEURVES-FRZN	FRUIT - DRIED
	PREPARED FOODS-FROZEN	JAMS, JELLIES, SPREADS
	PUDDING, DESSERTS-DAIRY	JUICE, DRINKS - CANNED, BOTTLED
	SNACKS, SPREADS, DIPS-DAIRY	NUTS
	UNPREP MEAT/POULTRY/SEAFOOD-FRZN	PACKAGED MILK AND MODIFIERS
	VEGETABLES-FROZEN	PASTA
	YOGURT	PICKLES, OLIVES, AND RELISH
		PREPARED FOOD-DRY MIXES
		PREPARED FOOD-READY-TO-SERVE
		SALAD DRESSINGS, MAYO, TOPPINGS
		SEAFOOD - CANNED
		SHORTENING, OIL
		SNACKS
		SOFT DRINKS-NON-CARBONATED
		SOUP
		SPICES, SEASONING, EXTRACTS
		SUGAR, SWEETENERS
		TABLE SYRUPS, MOLASSES
		TEA
		VEGETABLES - CANNED
		VEGETABLES AND GRAINS - DRIED

Note: [a]NielsenIQ non-dry food categories span five departments including dairy, deli, fresh produce, frozen foods, and packaged meat. We exclude ICE and YEAST as our elasticity estimation records show that their UPC-level products are not commonly shared by neighboring markets such that there are missing elasticity estimates of these categories for many markets. These two categories have limited revenue shares within each store though. [b]NielsenIQ dry food refers to food categories in the dry grocery department; we exclude GUM and PET FOOD that are not of research interest in this paper. The bold food categories are the ones that IRI and NielsenIQ share the same names, although NielsenIQ have more varieties or larger numbers of UPCs within each of these five common food categories.

C Data construction for market-year-specific factors

Several of our market-specific factor measures in the raw data, e.g., house prices, are nominal. We use the standard GDP deflator, measured at the national level, to take out the trend

Table B.5: Summary Statistics of Food Categories.

	IRI categories[a]			NielsenIQ categories[b]		
	mean	p10	p90	mean	p10	p90
obs per year (million)	5.9	1.6	12.2	29.8	7.4	63.4
No. of UPCs per year	2083.4	302.0	4771.6	4009.6	599.6	9810.9
No. of UPCs per year-market	546.0	122.4	1068.0	942.0	173.2	1975.1

Note: [a]IRI has 16 food categories. [b]NielsenIQ has 58 food categories.

Table B.6: Summary Statistics of the Equally-weighted Demand Elasticity Estimates.

	IRI sample[a]			NielsenIQ sample[b]		
	mean	p10	p90	mean	p10	p90
point estimate	2.46	1.68	3.28	1.62[c]	0.90	2.42
standard error	0.19	0.09	0.33	0.13	0.05	0.26
t statistics	16.58	7.53	27.23	18.55	6.22	33.30
obs in an estimation (thousands)	120.1	19.7	258.5	1147.5	67.6	2874.3
Cragg-Donald F statistics	133.1	15.0	276.0	1375.2	36.4	2664.9

Note: [a]The IRI sample has 3,915 elasticity estimates. [b]The NielsenIQ sample has 22,214 elasticity estimates. These samples include elasticity estimates that are obtained under the strong price IV and exclude those that have negative values. For each sample, we also trim off the upper and lower 1% of the remaining elasticity estimates in each year as we do for the revenue-weighted elasticity estimates. [c]The NielsenIQ sample has lower elasticity estimates than the IRI sample. The downward trend in demand elasticities over time largely contributes to the lower elasticity estimates in the NielsenIQ sample of more recent years. The NielsenIQ sample also has a small proportion of price-inelastic goods such as baby food, spices, seasoning, extracts and eggs.

increases in prices. We reset its base year to 2000, right before our sample period 2001-2022.

Real GDP per capita We deflate the county-level nominal GDP to the base year 2000 using the GDP deflator. Then, we divide it by population to obtain the real GDP per capita for every county in each year. Finally, we take logarithm and then average using county-level populations as weights to obtain the market-level counterpart.

Unemployment rate The raw data is in percentage. We average them using county-level populations as weights to obtain the market-level measurement for each year.

Cumulative change in real housing price We start with the county-level housing price index that is calibrated using appraisal values and sales prices for mortgages bought or guaranteed by Fannie Mae and Freddie Mac (Bogin et al., 2019). We choose the version with the base year 2000, the same base year for our GDP deflator, and deflate it using the GDP deflator. Then, we take differences relative to the base year for every county in each year, namely the county-level cumulative percentage changes relative to the base year. Finally, we average them using county-level populations as weights to obtain the market-level cumulative percentage change in real housing prices.

Population We aggregate annual county populations at the market level before taking logarithm.

Table B.7: The Factor Regression Results for Estimated Elasticities.

explanatory variables[a]	dependent variable: *demand elasticity estimate*[b] 2001-2022		2004-2022	
	short-term (estimated)	long-run[c] (imputed)	short-term (estimated)	long-run (imputed)
lagged elasticity	0.51***		0.52***	
	(0.02)		(0.02)	
real interest rate	3.91***	7.54***	6.18***	6.23***
	(1.19)	(1.74)	(0.98)	(1.36)
real GDP per capita (pre-COVID)	-0.34	-1.41***	-0.91**	-1.45***
	(0.42)	(0.37)	(0.33)	(0.32)
unemployment rate (pre-COVID)	2.87***	1.85**	1.56***	3.99***
	(0.80)	(0.92)	(0.46)	(0.98)
cum. perct change in real housing price	-0.24***	-0.34***	-0.24***	-0.00
	(0.07)	(0.09)	(0.05)	(0.10)
population	-0.27	-0.32	-0.42*	-0.60
	(0.31)	(0.64)	(0.23)	(0.45)
No. of UPCs per category	0.17***	0.12**	0.17***	0.18***
	(0.02)	(0.06)	(0.02)	(0.06)
grocery establishments per 1k residents	0.37	0.62		
	(0.26)	(0.52)		
No. of employees per grocery store	0.01	0.11		
	(0.03)	(0.07)		
retailer HHI			-0.25*	-0.36***
			(0.12)	(0.09)
post-COVID controls[d]				
interaction with real GDP per capita	1.28		2.24**	
	(0.97)		(1.04)	
interaction with unemployment rate	-1.84		1.37	
	(1.66)		(1.35)	
interaction with market dummies	YES		YES	
one lag of all variables (no dummies)	YES		YES	
fixed effects of markets and goods	YES		YES	
$adj.R^2$	0.71		0.70	
within R^2	0.44		0.47	
N	23,322		22,527	

Note: [a]Explanatory variables are logged except for real interest rate, unemployment rate, and cumulative percent change in real housing price. See details in the main text and Appendix C. [b]Elasticity estimates are market-category-year-specific and in levels. We shifted IRI elasticity estimates down by a constant such that the IRI and NielsenIQ samples have the same elasticity mean for the overlapping period of 2006-2012. [c]Given the distributed lag model, we impute the long-run effects for each factor and estimate standard errors using the Delta method. [d]Pandemic controls have two parts: one captures level shifts and the other captures interactions with GDP and unemployment. All the models above are estimated by OLS using market-category-specific revenue shares as weights within each year. Standard errors, clustered at the market level, are listed in the parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: The Factor Regression Results for Implied Markups with a Linear Year Trend.

explanatory variables[a]	dependent variable: *implied markup*[b] 2001-2022		2004-2022	
	short-term (estimated)	long-run[c] (imputed)	short-term (estimated)	long-run (imputed)
lagged implied markup	0.57***		0.57***	
	(0.02)		(0.02)	
real interest rate	-6.10***	-8.38***	-7.74***	-10.10***
	(1.02)	(2.26)	(0.88)	(2.37)
real GDP per capita (pre-COVID)	1.09**	1.89**	1.68***	2.27***
	(0.49)	(0.71)	(0.45)	(0.68)
unemployment rate (pre-COVID)	0.20	-1.28	0.82	-4.00***
	(0.62)	(1.00)	(0.64)	(1.27)
cum. perct change in real housing price	0.38***	0.34***	0.31***	0.00
	(0.07)	(0.12)	(0.07)	(0.16)
population	-0.20	0.30	0.24	1.24
	(0.26)	(0.74)	(0.30)	(0.75)
No. of UPCs per category	-0.17***	-0.27***	-0.18***	-0.33***
	(0.04)	(0.07)	(0.03)	(0.09)
grocery establishments per 1k residents	-0.15	-0.07		
	(0.25)	(0.40)		
No. of employees per grocery store	-0.03	-0.13		
	(0.03)	(0.10)		
retailer HHI			0.24**	0.23**
			(0.10)	(0.11)
post-COVID controls[d]				
interaction with real GDP per capita	-5.62***		-6.44***	
	(1.82)		(1.82)	
interaction with unemployment rate	-10.53***		-12.16***	
	(3.01)		(3.14)	
interaction with market dummies	YES		YES	
linear year trend	0.008		-0.001	
	(0.005)		(0.005)	
one lag of all variables (no dummies)	YES		YES	
fixed effects of markets and goods	YES		YES	
$adj.R^2$	0.66		0.65	
within R^2	0.48		0.48	
N	23,322		22,527	

Note: [a]Explanatory variables are logged except for real interest rate, unemployment rate, and cumulative percent change in real housing price. See details in the main text and Appendix C. [b]We obtain implied markups from equation (3); these markups are market-category-year-specific and in levels. In the calculation, we shifted IRI elasticity estimates down by a constant such that the IRI and NielsenIQ samples have the same markup mean for the overlapping period of 2006-2012. [c]Given the distributed lag model, we impute the long-run effects for each factor and estimate standard errors using the Delta method. [d]Pandemic controls have two parts: one captures heterogeneous level shifts across markets and the other captures interactions with GDP and unemployment. All the models above are estimated by OLS using market-category-specific revenue shares as weights within each year. Standard errors, clustered at the market level, are listed in the parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

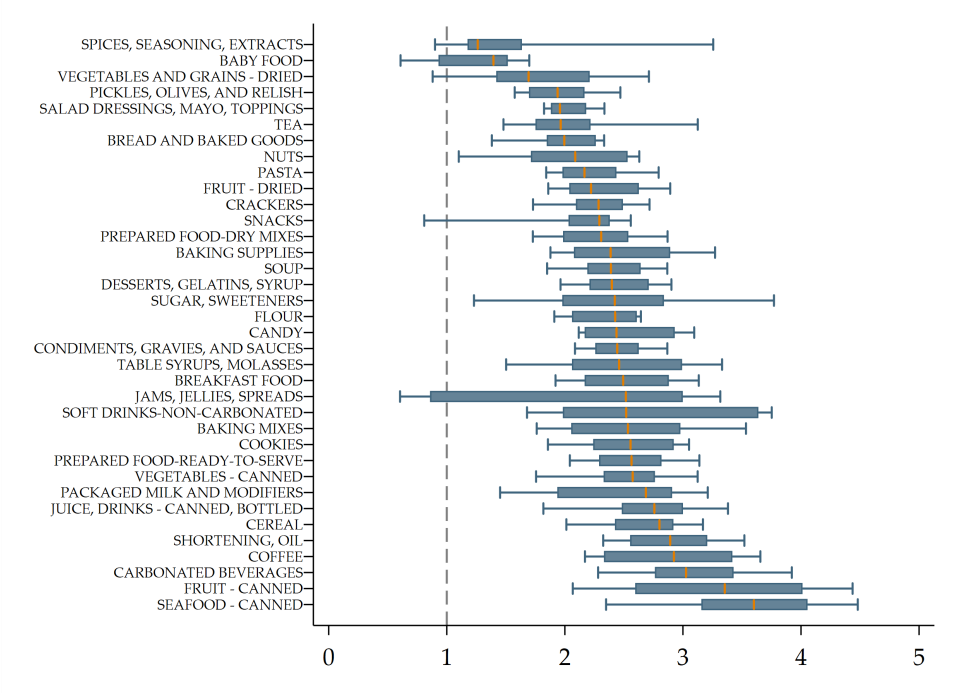


Figure B.1: NielsenIQ Demand Elasticity Estimates by Dry Food Categories in 2010.

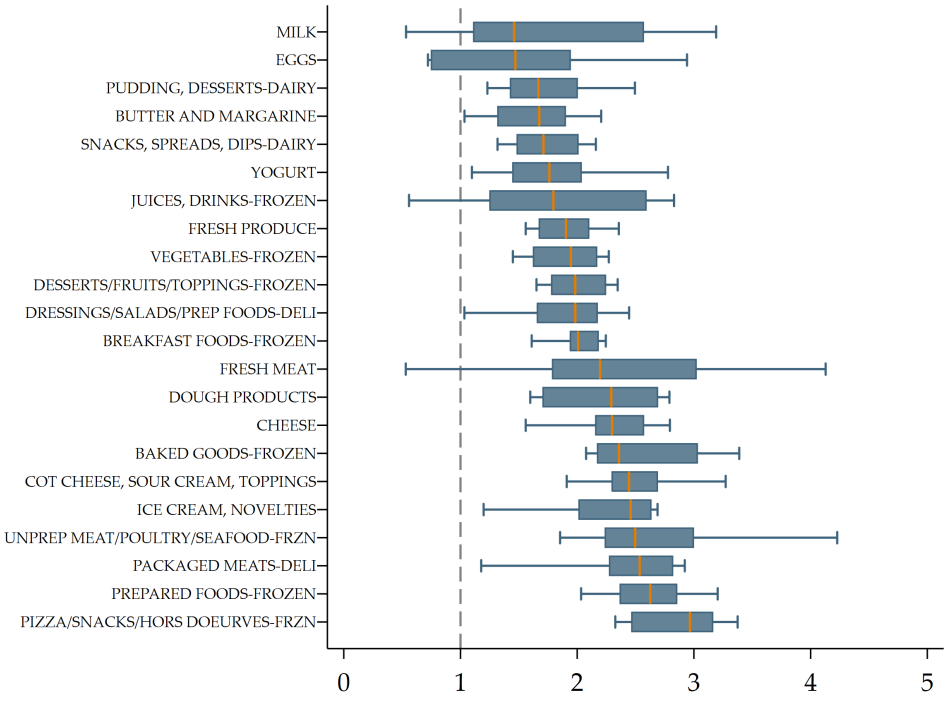


Figure B.2: NielsenIQ Demand Elasticity Estimates by Non-dry Food Categories in 2010.

Grocery establishments per capita For each county, we divide the number of grocery stores,

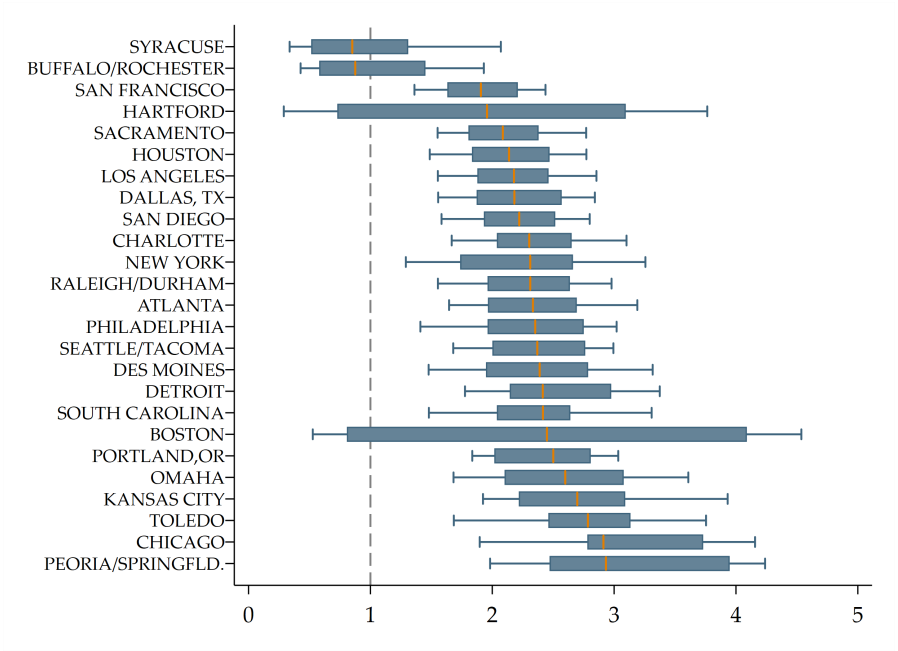


Figure B.3: NielsenIQ Demand Elasticity Estimates by Markets in 2010.

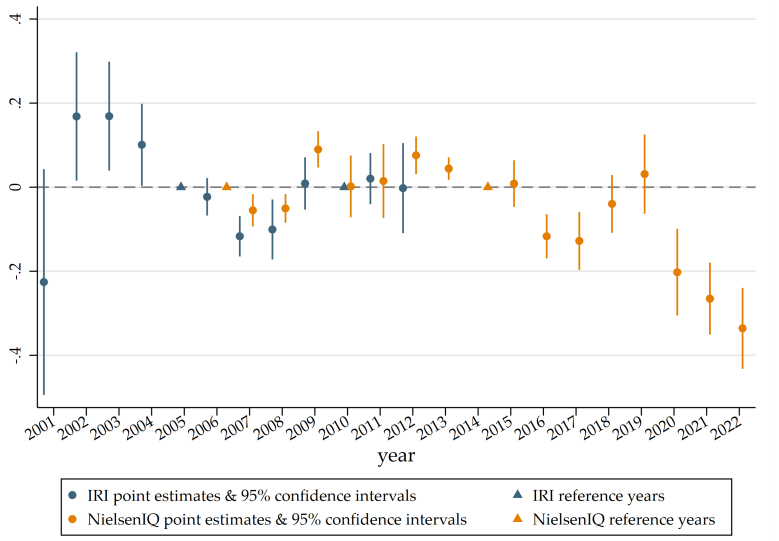


Figure B.4: Cyclical Variations of IRI and NielsenIQ Demand Elasticity Estimates.

Note: To obtain the cyclical variation in the IRI sample, we run demand elasticity estimates on year dummies and a linear year trend with market-good-specific fixed effects, using market-good-specific revenues as weights within each year. We run the same regression to obtain the cyclical variation of demand elasticity estimates in the NielsenIQ sample. For the IRI sample, we drop year dummies for 2005 and 2010 such that the cyclical variation around the linear trend has a statistically zero sum during 2001-2012. For the NielsenIQ sample, we drop year dummies for 2006 and 2014 such that the cyclical variation around the linear trend has a statistically zero sum during 2006-2019. The associated coefficients of year dummies and their 95% confidence intervals are plotted above.

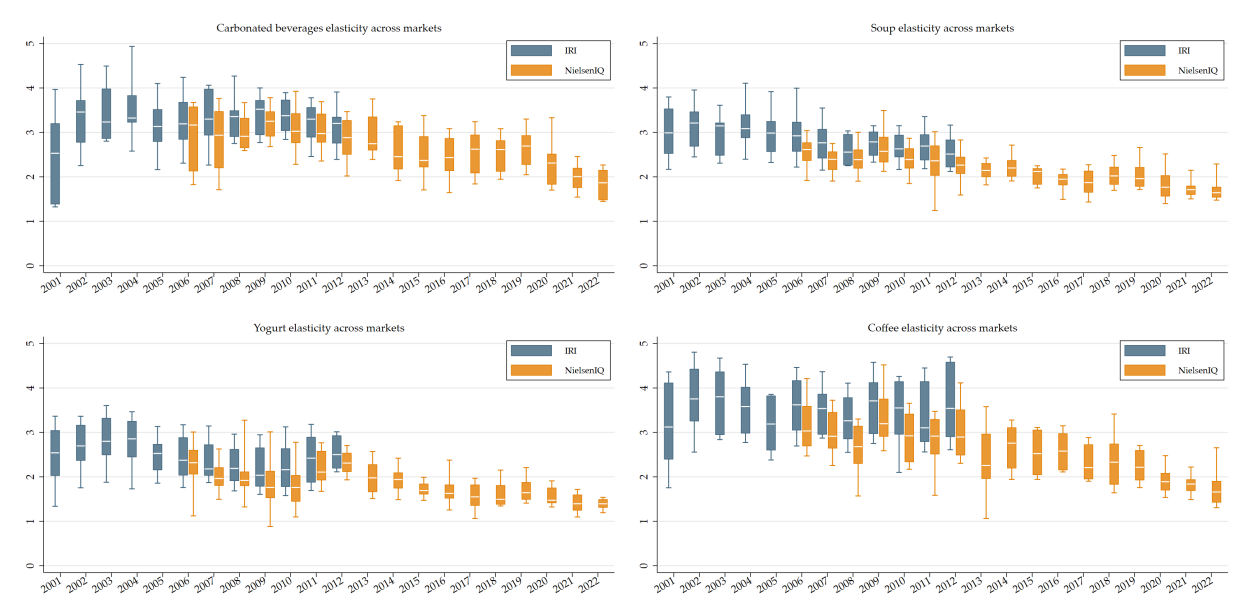


Figure B.5: Demand Elasticity Estimates of the Four Common Food Categories.

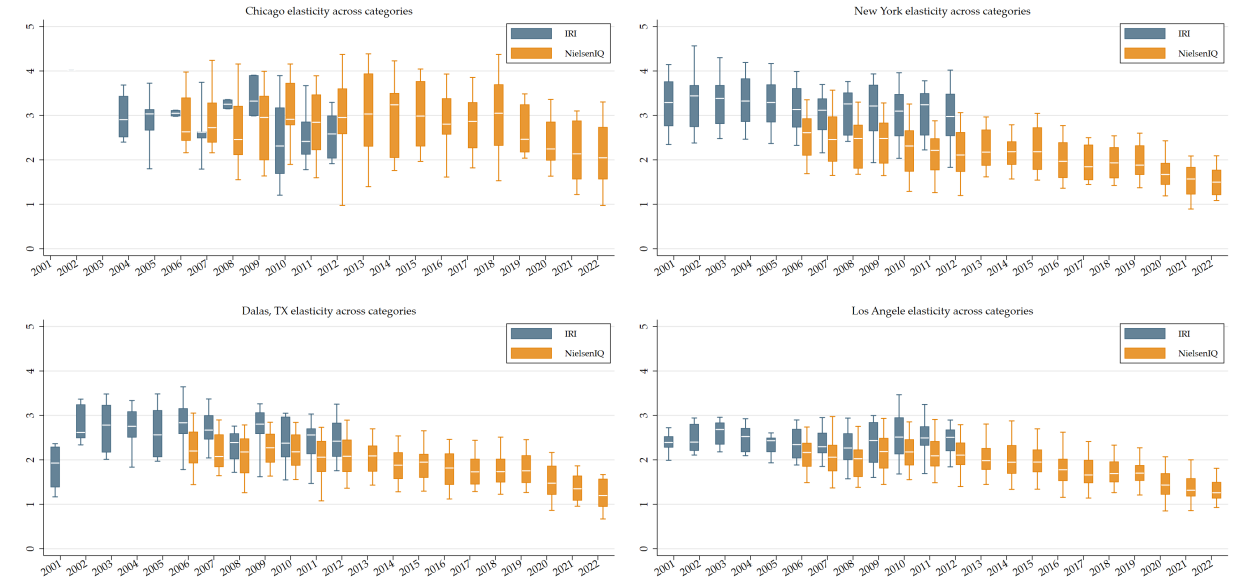


Figure B.6: Demand Elasticity Estimates in the Four Largest Regional Markets.

reported in the County Business Patterns from the Census Bureau, by its population and then times 1,000 to obtain the number of grocery stores per 1k residents across years. Then, we take logarithm before we average them into the market level using county-level populations as weights. This variable captures the availability of grocery stores.

No. of employees per grocery store For each county and year, we divide the total number of

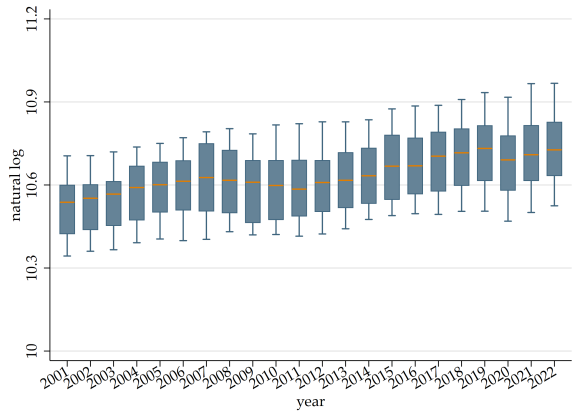
employees by the total number of grocery stores based on the County Business Patterns data from the Census Bureau. Then, we take logarithm before we average them into the market level using county-level populations as weights. This variable captures sizes of grocery stores.

Table C.1 below shows the public sources of the original data while Figure C.1 below gives their graphical representations. Other factors, including real interest rate, retailer-HHI, and UPCs per category, have been explained in Section 4.3. Figure C.2 below shows their dynamics over the years.

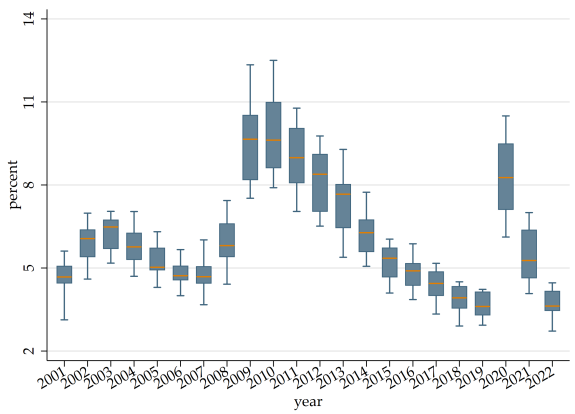
Table C.1: Public Sources of the Raw Data.

variable	level	source[a]
GDP deflator	national	Federal Reserve Bank of St. Louis
GDP	county	Federal Reserve Bank of St. Louis
housing price index	county	Federal Housing Finance Agency
unemployment rate	county	U.S. Bureau of Labor Statistics
population	county	U.S. Census Bureau
establishments of grocery stores	county	U.S. Census Bureau
No. of employees per grocery store	county	U.S. Census Bureau

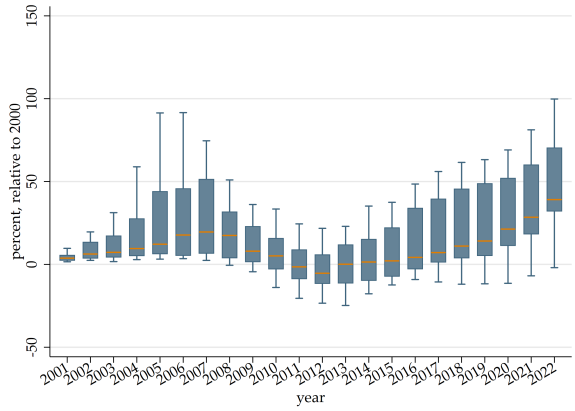
Note: [a]Here are website links below for all the data listed above.
(1) GDP deflator: "https://fred.stlouisfed.org/series/USAGDPDEFAISMEI".
(2) GDP: "https://fred.stlouisfed.org/release/tables?rid=397eid=1054597".
(3) housing price index: "https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx".
(4) unemployment rate & population employed: "https://www.bls.gov/lau/tables.htmcntyaa".
(5) population: "https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates.html".
(6) establishments of grocery stores: "https://www.census.gov/programs-surveys/cbp/data/datasets.html".



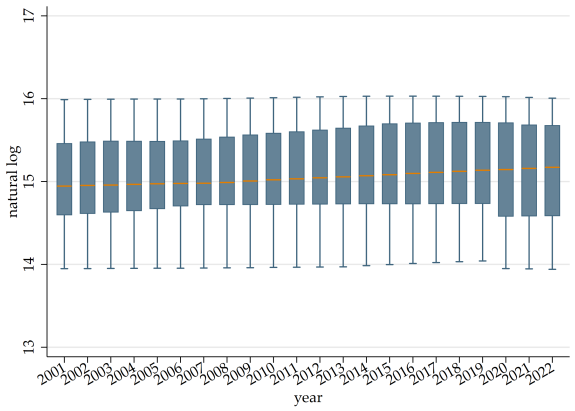
(a) real GDP per capita



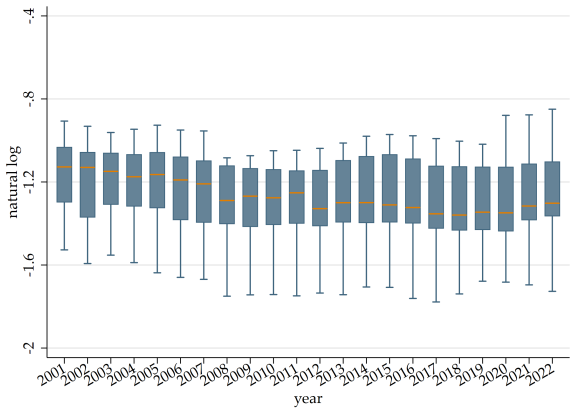
(b) unemployment rate



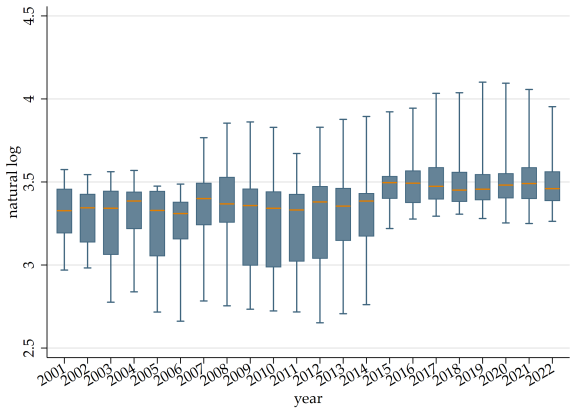
(c) cumulative change in real housing price



(d) population



(e) No. of grocery stores per 1k residents



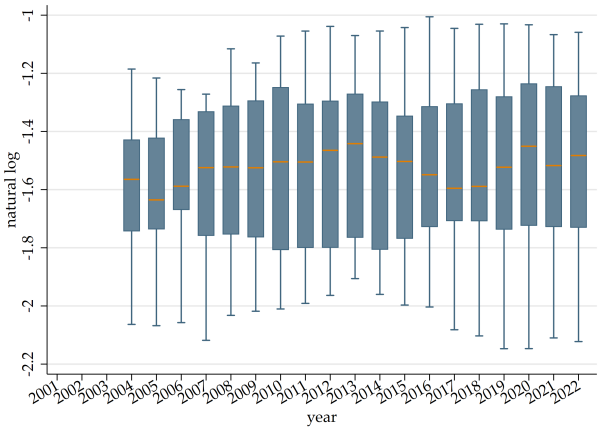
(f) No. of employees per grocery stores

Figure C.1: Cross-sectional and Time Variations of Market-specific Factors.

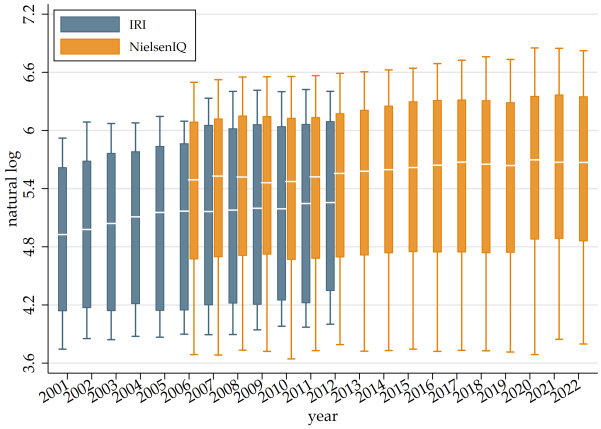
Note: See the text above for the constructions and descriptions of these market-level factors.



(a) One-year real interest rate



(b) Retailer-HHI



(c) UPCs per category

Figure C.2: Cross-sectional and Time Variations of Other Factors.

Note: The national one-year real interest rate is obtained from FRED. The market-level retailer-HHI is calculated from the NielsenIQ consumer panel that starts from 2004. We calculate the revenue-weighted average UPCs per category across stores from both IRI and NielsenIQ retail scanner data sets. This factor is at the market-category level. For the latter two factors, we show boxplots as in Figure C.1.

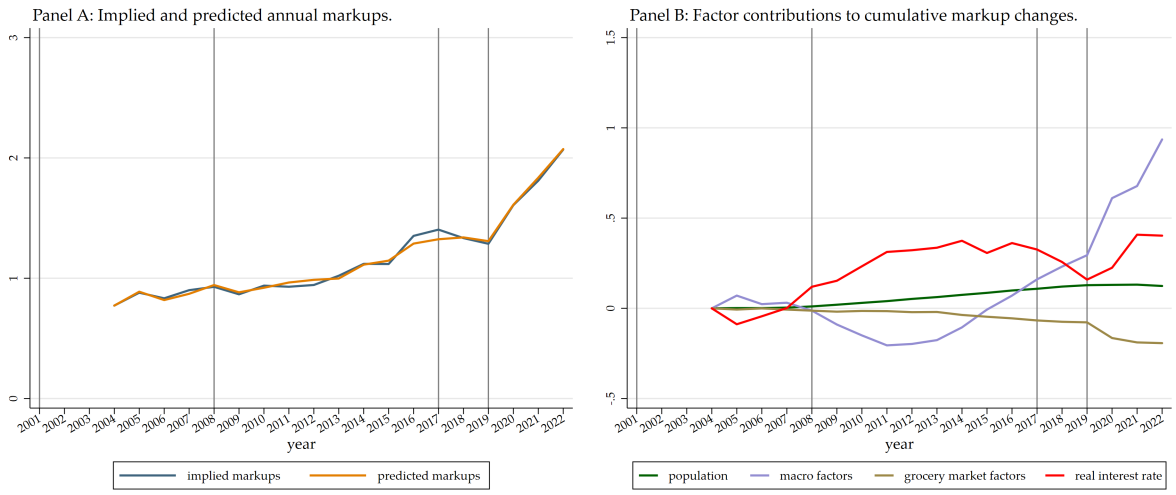


Figure C.3: The Revenue-weighted Annual Markups for 2004-2022.

Note: These results are based on the sample of 2004-2022 with retailer-HHI. As Figure 9, we first obtain predicted market-category-year-specific markups based on the estimated regression model (5). We then calculate revenue-weighted annual markups within the IRI and NielsenIQ samples, respectively. For the overlapping period of 2006-2012, we average annual markups using the IRI and NielsenIQ sample size shares as weights. We follow the same procedures for the annual markups implied from elasticity estimates. To calculate factor contributions, we first set all factors at their initial values to obtain predicted annual markups under self-dynamics, which are used as the baseline predicted annual markups. We then calculate the cumulative contributions of each group of factors by taking the cumulative differences between the factor-predicted annual markups and the baseline predicted annual markups. We group our eight factors into four groups: (1) real GDP per capita, unemployment rate, and cumulative changes in real housing prices combined as macro factors together with pandemic effects; (2) the number of grocery stores per capita, the number of employees per grocery store, and the number of UPCs per product category combined as grocery market factors; (3) population; and (4) real interest rate.