The Response of Retail Prices and Markups to Cyclical Demand Shocks

Tobias Renkin*

July 2017

Abstract

I study the response of markups and prices of products sold in US grocery stores to cyclical demand shocks between 2006 and 2010. I first show that under plausible assumptions on the production technology of multiproduct retailers, I can identify causal effects on markups using a combination of fixed effects and appropriate demand shifters. I then use variation in products' income elasticity of demand to construct cyclical demand shifters that are plausibly unrelated to cyclical supply shocks. I find that in response to a positive 10% demand shock, markups decrease by up to 5pp. My results support theories of countercyclical product level markups as proposed, for example, in Ravn et al. (2006, 2008).

*University of Zurich, tobias.renkin@gmail.com

1 Introduction

Macroeconomists have long been interested in the cyclical behavior of markups. This interest stems from the fact that in a wide range of business cycle models, amplification of productivity or demand shocks happens through counter-cyclical variation in markups. Moreover, expansionary fiscal and monetary policy typically work through channels that end up reducing markups. The idea of counter-cyclical markups is somewhat counter-intuitive at first, but is generated by a wide range of popular business-cycle models. However, due to the difficulties in measuring marginal cost and markups, the empirical evidence on the cyclical behavior of markups is limited.

In this paper, I study the movements of quantities, prices and markups of grocery products over the business cycle. While macroeconomists have extensively studied the developments of these variables in the aggregate, I exploit very detailed product level data on prices and quantities sold in different grocery stores in the US. Empirical work on markups faces one of the more challenging measurement issues in economics. Typically, econometricians observe prices and at best factor prices or a measure of average cost, and need to rely on assumptions to back out markups over marginal cost. I use a novel empirical strategy to address this measurement issue in the context of multi-product firms. Intuitively, I leverage variation in demand at the product level, while controlling for marginal cost at the firm level. A second challenge in studying the response of markups to demand shocks, is that demand and markups are determined simultaneously. I construct demand shifters based on product characteristics that serve as proxies for a products income elasticity of demand, together with local business cycles. I use these variables as instruments that shift demand but are independent of supply shocks. I show that these demand shifters are relevant instruments, and valid under plausible identification assumptions.

I derive such assumptions from a theoretical framework of multi-product retail production. The model has two important features. First, it features decreasing returns to scale at the store level, to incorporate the idea that capacity is fixed in the short run. However, product level marginal cost is still independent of product level output. This follows from the fact that multi-product firms can optimally reallocate fixed factors between the products they provide and thus equate the marginal productivity of input factors across products. Second, it assumes integrated supra-regional markets for all product-specific inputs, so that product-specific marginal cost does not correlate with local business cycles. Under these assumptions, marginal cost of multi-product firms can be decomposed in a location-specific component, a product-specific component, as well as productivity and markup shocks. This allows me to absorb potentially cyclical location- and product-specific components in a combination of controls and fixed effects, and rely on instruments that are independent of remaining supply shocks.

I find that that stores decrease markups in response to positive demand shocks. The magnitude of this effect is sizable, and my preferred estimate of the demand elasticity of markups is -0.5.

This implies that a 10% positive demand shock would decrease markups by 5 percentage points. The elasticities are imprecisely estimated and depend somewhat on the instruments I employ, but point estimates are consistently negative and significantly different from zero in most specifications. The imprecision of my estimates is due to a weak, but significant, first stage relationship between proxies for the income elasticity of demand and cyclical movements in quantities.

My findings speak to the theoretical literature featuring counter-cyclical markups. Most prominently, New Keynesian models strongly rely on counter-cyclical markups as an amplification mechanism. In these models, marginal cost is pro-cyclical, but prices are sticky. As a result, marginal cost fluctuations are not fully passed-through into prices, and the markup is countercyclical. The literature has also proposed models where markups behave in a counter-cyclical way even when prices are flexible. Rotemberg and Woodford (1992) suggest that competition between firms increases when output is high. In their model, it becomes more attractive to undercut a collusive price and capture the whole market when demand is high. As a result, the price that can be sustained in an equilibrium of "implicit" collusion depends negatively on demand. Chevalier and Scharfstein (1996) and Gilchrist et al. (2014) suggest that during recessions financially constrained firms sacrifice investments in market share in favor of higher markups to increase short-run profits and stay afloat. My empirical approach is most closely related to the work of Ravn et al. (2006, 2008) who suggest that markups are countercyclical because the product-specific price elasticity of demand decreases in output. As a result, markups decline when demand is high. This result follows from the assumption that demand consists of a fixed, price-inelastic, and a time-variable price-elastic component. The price-inelastic component of demand can be modeled either as subsistence consumption (Ravn et al., 2008) or as slowly adjusting consumption habits (Ravn et al., 2006). When output is high, the fixed price-inelastic component of demand is less important, and markups rise as a result.

My results support the relevance of theories of countercyclical markups, in which the price elasticity of demand is directly linked to product level output, and decreases when output is high, as in Ravn et al. (2006, 2008). Since I consider variation in output and markups conditional on marginal cost, my results do not provide evidence for or against countercyclical markups due to sticky prices and incomplete pass-through of changes in marginal cost, as in New Keynesian models. Moreover, since I consider variation between products sold in the same location by the same firms, my approach also doesn't directly relate to theories that propose countercyclical markups as a result of time-varying *firm-level* trade-offs between short run profits and market share, such as Rotemberg and Woodford (1992), Chevalier and Scharfstein (1996) or Gilchrist et al. (2014).

This paper proceeds as follows. I first shortly describe the related empirical literature. I then describe a framework of retail supply, and derive identification assumptions. Next, I describe the data used to estimate the retail supply curve. I then present estimates of the demand elasticity of markups. The final section concludes.

2 Literature review

The existing empirical literature on the cyclical movements of markups has come to no clear conclusion. The literature has focused on studying the cyclical behavior of the "aggregate markup", which is the inverse of aggregate real marginal cost. Typically, strong assumptions are required for the concept of aggregate markups and cost to be meaningful, and additional assumptions to construct measures of both from available data.

The most prominent method is to measure real marginal cost using the labor input margin. Under the assumption of a Cobb-Douglas production function, the markup is inversely proportional to the aggregate labor share. Since the labor share ist strongly countercyclical, that would imply pro-cyclical markups. Bils (1987), Rotemberg and Woodford (1999) and Gali et al. (2007) all apply a number of adjustments to this raw markup measure. For example, they allow for fixed labor costs or frictions in substitution and adjustment of production factors. After these adjustments, aggregate markups are countercyclical. Other authors have conducted more indirect inference on the cyclicality of markups. For example, the incentive to build up inventories is driven by time-variation in the spread between cost of production and prices—the markup. Inventories can be observed more easily than marginal cost, and variation in markups can be backed out using data on inventories combined with a structural model of firm behavior. Bils and Kahn (2000) devise such a model and find that in order to match the cyclicality of the sales to inventory ratio, markups need to be countercyclical. In contrast, Ramey (1991) estimates a cost function for several US manufacturing industries and finds that firms operate in regions of declining marginal cost. This suggests that if prices are sticky, markups should increase when quantities go up, and that markups are consequently pro-cyclical. Finally, Hall (2014) proposes a similar indirect approach. He argues that advertisement is especially profitable during times when markups are high, and suggests that the pro-cyclical nature of advertisement expenditures implies that markups cannot be countercyclical. Overall, it is fair to say that the evidence on the cyclical behavior of aggregate markups is mixed.

A related literature has studied the behavior of prices using microdata on prices and quantities. Gilchrist et al. (2014) highlight the importance of financial frictions, and show that financially weak firms increased prices during the 2008-09 recession, while financially strong firms decreased their prices. However, they cannot separate the effect of demand shocks from other potential channels. Chevalier et al. (2003) study the price response to seasonal demand shocks around major holidays and find that prices decline in response to positive demand shocks. Gagnon and Lopez-Salido (2014) use demand shocks associated with strikes of neighboring stores and extreme weather events. They find that these events translate into large demand shocks for some stores, but do not move prices.

I contribute to the existing empirical literature in two ways. First, the macroeconomic literature relies on very strong assumptions to identify movements in aggregate markups, which are in themselves hard to interpret. The smaller empirical literature studying specific demand shocks using price microdata identifies credible effects, but it is difficult to tell whether the response to demand shocks associated with holidays, labor conflicts and extreme weather events generalizes to smaller and less extraordinary cyclical demand shocks.

3 Framework and identification

I use product characteristics as predictors for how demand for a given product varies over the business cycle. The resulting cyclical, product-specific demand shifters allow me to identify the response of prices and markups of individual products to cyclical demand shocks. Intuitively, the identifying assumption is that conditional on controls, these product characteristics are predictive of cyclical movements in demand, but independent of potential shocks to supply. In what follows, I derive explicit identification assumptions from a theoretical framework of retail supply. This framework serves two purposes. First, it illustrates the necessary controls to absorb potentially cyclical variation in marginal cost. Second, it helps to illustrate the exclusion restrictions on the instruments. The model has two key features. First, under some assumptions on multi-product retail stores' production technology, product-level marginal cost does not vary with product-level quantities, even though total store capacity is fixed in the short run. Second, under the assumption of an integrated wholesale market, marginal cost can be decomposed into a regional and a product-specific component, which can be absorbed in appropriate fixed effects and controls. I can thus keep marginal cost constant, and interpret causal effects of demand shocks on prices as effects on markups. My approach does not require me to specify a particular market structure or demand system. Beyond identifying a relevant demand shifter, it thus remains largely neutral on the exact relationship between demand and business cycles.

3.1 The aggregate local retail supply curve

Retail technology I consider grocery stores that sell many products. For each product, a grocery store buys merchandise in a national wholesale market and combines it with several local inputs such as labor or real estate. The final product of the store—retail output—should be seen as the physical merchandise itself, augmented with the service of providing easy shopping to local customers.

Stores are indexed by j and are situated in local market $s(j)$. I suppress time subscripts for readability for now. To provide retail output of product i, a store purchases X_{ij} units of merchandise at price R_i. Merchandise of type i can only be used to provide retail output of product i. In addition, the store employs L_j units of labor at wage $W_{s(j)}$. L_{ij} units of labor are devoted to the provision of retail output of product i. Labor can be freely adjusted and redirected to the provision of different products. Finally, the store has a fixed overall capacity K_j, which I normalize to 1. Capacity embodies real estate, shelf space, checkout counters and so on. The store devotes a share K_{ij} of capacity to the provision of product i. Overall capacity cannot be adjusted, but the existing capacity can be freely allocated over all products the store provides. Finally, retail output is affected by a product-store specific productivity shock a_{ij}. For each product, retail output is produced using the same Cobb-Douglas technology:

$$Q_{ij} = L_{ij}^{\alpha} X_{ij}^{\beta} K_{ij}^{1-\alpha-\beta} e^{-a_{ij}} \tag{1}$$

This production function features constant returns to scale at the product level. This implies that conditional on its overall output, the store can operate equally efficiently independent of whether inputs are divided between few or many different products. However, because overall capacity is fixed, returns to scale are decreasing in aggregate store output. If capacity could be adjusted, returns to scale would be constant even in the aggregate. In this sense, capacity is to be interpreted very broadly.

Retail cost function The store minimizes its total cost given factor prices and output $\{Q_{ij}\}$. The store's cost function is derived in detail in the Appendix .A.1 and given by:

$$C(\{Q_{ij}\}, \{R_i\}, W_{s(j)}) = \Omega W_{s(j)}^{\frac{\alpha}{\alpha+\beta}} \left(\sum_{i=0}^{I} Q_{ij} R_i^{\beta} e^{a_{ij}} \right)^{\frac{1}{\alpha+\beta}}$$

Ω is a constant that depends on α and β. The marginal cost of providing an additional unit of product i is given by:

$$MC_{ij} = W_{s(j)}^{\frac{\alpha}{\alpha+\beta}} R_i^{\beta} e^{a_{ij}} \frac{\Omega}{\alpha+\beta} \left(\sum_{i=0}^{I} Q_{ij} R_i^{\beta} e^{a_{ij}} \right)^{\frac{1-\alpha-\beta}{\alpha+\beta}}$$

$Q_j \equiv \left(\sum_{i=0}^{I} Q_{ij} R_i^{\beta} e^{a_{ij}} \right)$ is a measure of total store output, weighted by product-specific cost, and reflects the fact that at higher output, capacity becomes scarce, adjustable inputs become less productive, and marginal cost goes up. If $\alpha + \beta = 1$, capacity is irrelevant for production and marginal cost does not depend on total store output. However, even though capacity is fixed, marginal cost of product i depends only on total store output, but not on output of product i. This is a result of optimal allocation of existing capacity between products, which works to equalize the marginal product of shared inputs between different products.

Retail markups and local supply curve Firms maximize profits for each product separately. Markups are given by $\mathcal{M} = \sigma(Q_{ij}, e^{\varepsilon_{ij}})/(\sigma(Q_{ij}, e^{\varepsilon_{ij}}) - 1)$. $\sigma(Q_{ij}, e^{\varepsilon_{ij}})$ denotes the price elasticity of demand, which may depend on quantities and a markup shock that is independent of quantities.

I now approximate the optimal price around a long-run equilibrium P_{ij}^0, Q_{ij}^0, ε_{ij}^0. The point of approximation is irrelevant and will be absorbed by appropriate fixed effects in my empirical analysis. For ease of exposition I do not change the notation, but it should be clear that the following is a linear first-order approximation. I assume that $\Gamma^q \equiv \partial\mathcal{M}(Q_{ij}^0, 1)/\partial Q_{ij}$ and $\Gamma^\varepsilon \equiv \partial\mathcal{M}(Q_{ij}^0, 1)/\partial e^{\varepsilon_{ij}}$ is the same across all stores and products.

$$p_{ij} = \Gamma^q q_{ij} + \Gamma^\varepsilon \varepsilon_{ij} + mc_{ij}$$

The fact that Γ^k does not vary between stores and products, allows me to aggregate deviations across different stores to derive the local supply curve in location s. The aggregate price of product i depends on aggregate local retail output of product i, the aggregate local markup shock $\overline{\varepsilon}_{is}$ and aggregate local product-specific marginal cost. To sum up, at any given time, the aggregate local supply curve is given by:

$$\overline{p}_{is} = \Gamma^q \overline{q}_{is} + \underbrace{\frac{\alpha}{\alpha+\beta} w_{s(j)} + \frac{1-\alpha-\beta}{\alpha+\beta}\overline{q}_s}_{\text{Location specific}} \underbrace{+\beta r_i}_{\substack{\text{Product}\\ \text{specific}}} + \overline{a}_{is} + \Gamma^\varepsilon \overline{\varepsilon}_{is} \tag{2}$$

The log-linear decomposition of marginal cost into a location-specific and a product-specific component is crucial for my empirical approach. Hence I shortly recapitulate the assumptions that are required for it. First, I assume that wholesale cost is the same for different stores, and does not depend on quantities. This rules out quantity discounts and price discrimination of wholesalers. In the United States, such practices are illegal under the Robinson-Patman act. Second, I assume that production technology is the same for all products a store provides. Given that my data includes only packaged consumer products, I view this assumption as unproblematic. Third, I assume that labor and capacity can be freely reallocated across products. I would argue that this reallocation happens largely by default—for example, there are no product-specific checkout counters or stock clerks. Finally, I assume that if capacity could be adjusted freely, returns to scale would be constant. This takes broad definition of capacity, and I view this assumption in line with a replication argument: if a store were to be copied in its entirety, and the copy placed next to the existing one, the two of them taken together would be equally productive as a single one.

3.2 Empirical model and demand shifters

Model of demand My main interest is in the response of markups to demand shocks that result from business cycle movements. Because the time dimension is central to the analysis of business cycles, I now introduce time subscripts to all variables. Let y_{st} denote the state of the local business cycle. I assume that there are some demand shifters that allow for the decomposition $\varepsilon_{ist} = \gamma_\varepsilon z_i y_{st}$

and that local demand for product i is consequently given by:

$$\overline{q}_{ist} = \gamma_\varepsilon z_i y_{st} + \gamma_p \overline{p}_{ist} + \overline{v}_{ist}$$

I can use $z_i y_{st}$ as a demand shifter in order to identify the response of markups to demand shocks using an IV estimator. I use various product characteristics z_i, all of which should be seen as proxies for the income elasticity of demand for product i and which will be discussed below. However, the validity of these instruments will depend critically on the appropriate controls to absorb potentially cyclical variation in marginal cost.

Empirical model Marginal cost may potentially depend on y_{st}, even if it is independent of product-level output. Hence, in order to identify the effects of a demand shock, I need to control for changes in marginal cost in my estimation. However, I don't observe important components of marginal cost. Therefore, I rely on the fact that variation in log marginal cost can be linearly decomposed into location-time-specific and product-time-specific factors under the assumptions on technology outlined in the previous section. I do not observe wholesale prices r_{it}, but I can absorb them in product-time fixed effects. Moreover, I do not observe cost-weighted store output $\overline{q}_{st}$. In my baseline estimation, I proxy for q_{st} with local business cycle conditions y_{st}. Alternatively, I can absorb q_{st} in location-time fixed effects, which I will do in robustness checks. Both strategies imply that I identify the response to demand shocks using only within-location-time variation in demand between different products. My estimation can be summarized as follows:

$$\overline{p}_{ist} = \Gamma^q \overline{q}_{ist} + \phi^w w_{st} + \phi^y y_{st} + \phi^t_{it} + \phi^s_{is} + \overline{a}_{ist} + \overline{\varepsilon}_{ist} \qquad \text{Supply curve} \qquad (3)$$

$$\overline{q}_{ist} = \gamma_\varepsilon z_i y_{st} + \gamma_p \overline{p}_{ist} + \overline{v}_{ist} \qquad \text{First stage} \qquad (4)$$

Identifying assumptions I now discuss the key assumptions necessary to estimate Γ^q. First, $z_i y_{st}$ needs to be a relevant instrument. This requires that z_i is a good predictor of how $\overline{q}_{ist}$ moves over business cycles. Second, $z_i y_{st}$ needs to be a valid instrument. This is the case if it is independent of markup and productivity shocks conditional on controls, i.e. $\mathbb{E}(z_i y_{st} \overline{a}_{ist} | controls) = 0$ and $\mathbb{E}(z_i y_{st} \overline{\varepsilon}_{ist} | controls) = 0$. The productivity and markup shocks $\overline{a}_{ist}$ and ε_{ist} can be seen as either as shocks in the true sense of the word, or error terms due to misspecification of the retail production technology and markup function. The main concern is that productivity and markup shocks may correlate with local business cycles, or that they correlate with product characteristics z_i. For example, the two processes could be given by:

$$\overline{a}_{ist} = y_{st}(1 + \psi_i) + z_i + \chi_{ist}$$
$$\overline{\varepsilon}_{ist} = y_{st}(1 + \eta_i) + z_i + \pi_{ist}$$

Where $\chi_{ist} \sim N(0, \sigma^2_\chi)$ and $\pi_{ist} \sim N(0, \sigma^2_\pi)$. The fact that productivity and markup shocks correlate one-to-one with y_{st} and z_i does not affect identification because my regressions include y_{st} and product fixed effects as controls. However, the instrument is invalid if the cyclical variation in product-specific demand driven by z_i is correlated with cyclical variation in product-specific productivity or markups, i.e. $\mathbb{E}(z_i\gamma_i|controls) \neq 0$ or $\mathbb{E}(z_i\eta_i|controls) \neq 0$. A sufficient condition for identification is that demand may vary in a way that is related to product characteristics, but conditional on marginal cost and demand, markup and productivity shocks occur at the store level but affect all products the same way.

Instruments A large literature studying consumer demand has shown that different goods exhibit widely varying (household) income elasticities of (household) demand (see for example Banks et al., 1997; Lewbel and Pendakur, 2009). This literature has estimated how the expenditure share of different goods varies with household incomes—the well-known "Engel curve". Some goods are necessities—their income elasticity is lower than one—and others are luxuries—their income elasticity is above one. I construct different instruments based on this idea, all of which should capture differences in the income elasticity of demand.

My approach departs from the previous literature studying consumer demand in two aspects. First, I do not look at goods in the sense that the demand systems literature typically defines them—i.e. food, transportation, etc.— but rather products within narrowly defined categories. Thus, I rely on the fact that even within a good such as food, demand for different varieties will vary with income and is not entirely determined by—for example—differences in preferences. Rather, richer households may consume, for example, higher quality products. Second, the ideal instrument would be the aggregate income elasticity of market demand, rather than the household income elasticity of household demand. The relationship between the two depends on the shape of Engel curves and the local distribution of incomes. The two will coincide only if an Engel curve is linear, or if there is no income inequality between households.

My first instrument is simply a product dummy, i.e. $z_i = 1_i$. Using this approach, the first stage of the TSLS estimator will estimate the aggregate income elasticity of market demand. This approach is the most general—it uses all variation in quantities that correlates with business cycles—but it leads to some econometric issues. Because I have to estimate one parameter for each product in the first stage, the number of instruments becomes very large. Overfitting the first stage regression is a serious worry, especially if the first stage relationship is weak. It is well known that the TSLS estimator with many weak instruments may be substantially biased toward OLS even in large samples, and that standard inference is unreliable (see for example Staiger and Stock, 1997; Stock et al., 2002). For this reason, I use two different product characteristics z_i, that should be seen as a proxy for the household income elasticity of demand. The second instrument is the average consumption-weighted income of households consuming a product. As the expenditure share of luxury products—with high income elasticities—increases with income, consumers of luxury products should on average be richer than consumers of necessities. The third instrument is the "expensiveness" of a product compared to other products in the same category (a category will correspond to, for example, frozen pizza). This relative price should be seen as a measure of quality. Bils and Klenow (2001) show that richer households systematically consume higher quality products, suggesting that higher quality products have higher household income elasticities of demand.

4 Data

I use a dataset containing unique product code (UPC) level revenues and quantities of about 2,377 US supermarkets described in Bronnenberg et al. (2008). The data is collected by market research firm IRI and covers the period 2001 to 2011. Overall, the data contains about 120,000 unique products in 31 categories. All products are standardized consumer packaged goods. Examples of categories are beer, frozen pizza or shampoo. Supermarkets are located in one of 530 counties in 41 states, and belong to one of about 90 retail brands. This dataset has been used in a number of economics papers, for example Coibon et al. (2015) or Nakamura et al. (2011).

Product definition In the raw data, products are defined and identified by their Unique Product Code. This is the finest conceivable level of disaggregation. An example for a unique product is a 12 oz. can of Coca-Cola Classic. Coca Cola Classic comes in many other sizes and containers, each of which is assigned its own UPC. In my empirical analysis, I bundle UPCs together to a slightly more coarse definition of product. In particular, I combine UPCs of the same vendor and brand to one single product. In most cases—such as Coca Cola Classic—this combines different package sizes. However, Coca Cola Classic and Coke Zero are treated as different brands, and hence will still be considered as distinct products. For other product categories, such as yogurt, this will also combine different flavors of yogurt, such as Dannon Vanilla, Dannon Strawberry and so on. The data comes with a volume or weight measure for each unique product, so that I can normalize prices and quantities by the volume or weight of each product before combining products. I bundle products together for two reasons. First, this allows me to reduce the number of parameters that need to be estimated when using product dummies as instruments. Second, it allows me to combine several products when computing the mean income of consumers from household panel data, and thereby serves to improve precision of this average.

Sample selection I conduct my analysis for the period from 2006 to 2010. This is in order to limit product and store turnover, and to focus on a period with large variation in cyclical local demand shocks. Furthermore, I restrict my analysis to a balanced panel of stores and products. I disregard all stores that enter the sample after Q1 2006 or exit before Q4 2010. Moreover, I drop stores that report sales for less than 90% of the quarters in between. I also drop products that are introduced after Q1 2006 or discontinued before Q4 2010 or sold in less than 90% of the quarters in between. Overall, this leaves me with a panel of 1289 stores and 1215 different products. For a part of my analysis, I will further restrict myself to a random draw of 500 different products due to computational constraints[1].

Local price index and quantities I first calculate weekly prices of each UPC u, in store$j$ and week w, by dividing revenues and volume adjusted quantities: $P_{ujw} = Revenue_{ujw}/(Quantities_{ujw} \cdot Volume_u)$. I then calculate quarterly prices for each product by calculating the unweighted mean price of all UPCs of product i, in store j and quarter t. Finally, I calculate the local price of product i in county s by aggregating across stores with fixed product-specific weights given by the average quarterly quantity of product i sold between 2006 and 2010. Quantities are constructed by summing (without weights) over quantities sold in all stores, UPCs and weeks observed for a product within a county in a quarter.

Construction of instruments I now discuss the construction of the instruments used in the paper. The first instrument is the "expensiveness" of a product. This measure is a proxy for the quality of a product, driven by the idea that higher quality products have higher income elasticities of demand. To construct expensiveness, I calculate the logarithm of the average quarterly price (volume adjusted) of a product in a county relative to all other products of the same category sold in the same county[2]. I then average this relative price over all counties and quarters between Q1 2006 and Q4 2010. The second instrument is the average income per member of all households that consume a product. This should also be seen as a proxy for the income elasticity of demand. The IRI data comes with a panel of 5000 households located in two cities: Eau Claire, WI and Pittsfield, MA. Households in the panel record their grocery consumption either at home using a barcode scanner, or directly at store checkouts using a specific IRI issued ID card, that works similar to loyalty cards used by many retail chains. IRI collects demographic data at the time the households enter the panel. For each product and year, I calculate the logarithm of the consumption-weighted average household income per member of consuming households, relative to average household income of all panel households in a given city and year. I then average this measure over all years. This results in a measure of how rich the average consumer in the panel is relative to all panel members.

Supplementary data I use quarterly data on labor earnings in NAICS sector 4451 "grocery stores" from the Quarterly Census of Employment and Wages (QCEW) as a proxy for wages. To construct average wages, I calculate the ratio of total labor earnings over total employment. Moreover, I use quarterly unemployment rates in a county as a measure of local business cycles y_{st}— since county level incomes are not available at quarterly frequency. The unemployment rates are provided in the Local Area Unemployment Statistics produced by the Bureau of Labor Statistics.

[1] The memory requirements of estimating a first stage with 1215 parameters are substantial.

[2] This includes all products and stores, i.e. also those that are not in the sample because they enter after 2006 or exit before 2010.

5 Results

5.1 Instrument relevance: heterogeneity in cyclical behavior of quantities

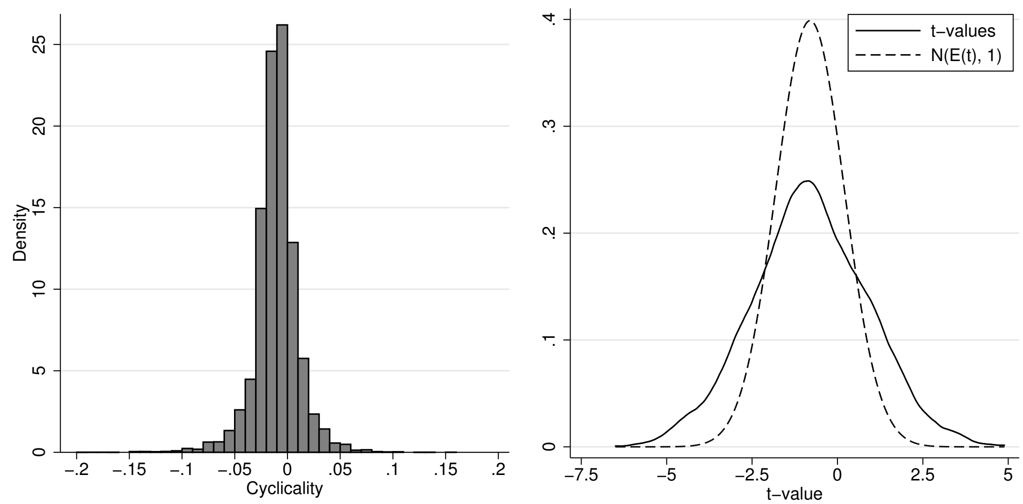
My identification strategy requires that demand for different products varies over business cycles in different ways, that are correlated with product characteristics z_i. These product characteristics are proxies for the income elasticity of demand. A large empirical literature has focused on the estimation of demand systems, and has shown that the income elasticity of demand varies widely between goods, and is in general not equal to one—preferences are very clearly not homothetic (see for example Banks et al., 1997; Lewbel and Pendakur, 2009). Hence, there is no ex-ante reason to expect that demand for different goods and products would vary with variations in aggregate income in the same way. However, I can test whether cyclical demand patterns for different products vary using my data. I run a regression of quantities on local unemployment:

$$q_{ist} = \phi_{is} + \phi_{it} + \gamma_w w_{st} + \gamma_i u_{st} + \varepsilon_{ist}$$

A negative γ_i indicates that product quantities move pro-cyclically, a positive γ_i indicates counter-cyclical behavior, and a γ_i of zero means that quantities are acyclical. Note that this specification is analogous to the first stage regression in equation 3. There is large variation in the estimated γ_i coefficients. This is illustrated in Figure 1a. 1b shows that this variation is not the result of sampling error by comparing the t-statistics of the γ_i estimates to a shifted standard normal distribution. The average product behaves very slightly pro-cyclically. The average elasticity is -0.015—i.e. a 10pp increase in local unemployment goes along with a decrease in sold quantities by 0.15%. However, there is a large fraction—30%—of products that behave counter-cyclically.

Table 1 illustrates some correlates of the estimated cyclical behavior. In columns (1) and (2) I consider product characteristics within all products or within categories. I find that more expensive and lower volume products are exhibit more pro-cyclical movements in quantities. The relationship of expensiveness and cyclical behavior is slightly nonlinear, with very expensive products being especially pro-cyclical. This result supports the idea that consumers react to downturns by buying cheaper staple products rather than expensive luxuries. In columns (3), (4), (5) and (6), I consider selected consumer characteristics, separately and in addition to the product characteristics. I find that products bought by households with higher incomes per household member are also more likely to exhibit pro-cyclical movements in quantities. This also supports the idea that the average consumer income can be used as a proxy for income elasticities of household demand, and that these predict aggregate income elasticities of market demand. Overall, the results confirm the idea that the cyclicality of products varies widely, and that expensiveness and household consumer income

Figure 1: Distribution of correlation of quantities and local unemployment over different products



(a) Distribution of correlation with local unemployment

(b) Distribution of t-values compared to shifted standard normal

Table 1: Correlates of cyclicality behavior of quantities

	(1)	(2)	(3)	(4)	(5)	(6)
Expensiveness	-0.006***	-0.005***			-0.005***	-0.005**
	(0.002)	(0.002)			(0.002)	(0.002)
Expensiveness squared	-0.003**	-0.004*			-0.004**	-0.005**
	(0.002)	(0.002)			(0.002)	(0.002)
Volume	0.002***	0.002***			0.002***	0.002***
	(0.000)	(0.000)			(0.001)	(0.000)
Private label	0.003	0.005			0.002	0.004
	(0.003)	(0.004)			(0.003)	(0.004)
HH income pc			-0.013***	-0.011**	-0.011***	-0.009*
			(0.004)	(0.005)	(0.004)	(0.005)
Owns property			0.004	0.003	0.005	0.004
			(0.003)	(0.004)	(0.003)	(0.004)
HH head some college			0.001	0.001	0.004	0.005
			(0.004)	(0.004)	(0.004)	(0.004)
Observations	1215	1215	1184	1184	1184	1184
r2	0.021	0.108	0.012	0.096	0.030	0.114
Category FE	NO	YES	NO	YES	NO	YES
Weights	1/var(γi)	1/var(γi)	1/var(γi)	1/var(γi)	1/var(γi)	1/var(γi)

Notes: Cyclicality is γ_i coefficient from the regression $q_{ist} = \phi_{is} + \phi_{it} + \gamma_w w_{st} + \gamma_i u_{st} + \varepsilon_{ist}$. Negative values of γ_i indicate pro-cyclical products. The distribution in Figure 1a and regressions in Table 1 are weighted by $1/var\gamma_i$. Table 1 reports robust SE in parenthesis, with * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

can be used as predictors of cyclical behavior of quantities. They both appear to be relevant instruments. However, both are not very strong instruments. The R^2 of specifications without category fixed effects lies between 0.012 and 0.03. As will become apparent later, this rather weak first stage relationship translates into large standard errors of the TSLS estimates.

5.2 Main results

Table 2 shows the main results of the empirical analysis. Column (1) presents the OLS estimates of equation 2. Prices and quantities are negatively correlated with a coefficient of -0.04. However, since both are determined simultaneously, there is no clear interpretation to this relationship. Prices are lower when unemployment is high, and correlate significantly with grocery store wages. Columns (2), (3) and (4) present the TSLS estimates using the three different instruments. Column (2) shows the estimates using product dummies as instruments. This specification estimates the unemployment elasticity of market demand in the first stage. The estimate for the demand elasticity of markups is -0.05 and precisely estimated. Yet, it should be interpreted with caution. Due to the high number of instruments in the first stage, it is likely that this estimator is biased toward OLS, and that standard inference is unreliable. Given that the product dummy TSLS estimate is more negative than OLS, and that it is biased toward OLS, it can be seen as an upper bound on the demand elasticity of markups.

Columns (2) and (3) show results for the expensiveness and consumer income instruments. Both estimates are more negative than estimates from OLS or TSLS with dummy instruments. This is consistent with the idea that the dummy instrument indeed leads to considerable bias toward OLS. The estimates of the demand elasticity of markups are -0.5 and -0.24 respectively and significant at 5% and 10% confidence levels. The first stage relationship is relatively weak, with F-statistics of around 5. This translates into large standard errors of the TSLS coefficients. However, just-identified TSLS is barely biased even with weak instruments, and inference typically reliable. All three specifications indicate that markups decrease with positive demand shocks, and my two preferred just-identified TSLS estimates indicate a substantial and economically very significant negative relationship.

Table 3 reports the same estimates for a more restrictive specification in which I include location-time fixed effects instead of controlling for wages and business cycles. The coefficients estimated using OLS and TSLS with product dummies are barely affected. The just-identified TSLS coefficients are both negative, and not significantly different from those reported in Table 2. However, the coefficient using the expensiveness instrument is no longer significantly different from 0.

Table 2: Main estimates of the demand elasticity of markups

Dep. var.: log Price	(1) OLS	(2) TSLS	(3) TSLS	(4) TSLS
log Quantity	-0.044***	-0.050***	-0.499**	-0.237*
	(0.001)	(0.003)	(0.232)	(0.142)
Unemployment	-0.002***	-0.002***	-0.007**	-0.004**
	(0.001)	(0.000)	(0.003)	(0.002)
log Wage	0.028***	0.031***	0.062*	0.043**
	(0.007)	(0.001)	(0.032)	(0.018)
Observations	4783401	1958571	4783401	4753813
Products	1215	500	1215	1193
Product time	YES	YES	YES	YES
Product county	YES	YES	YES	YES
County time	NO	NO	NO	NO
Instrument	NO	Product	EXP	INC
1st stage F-stat		9.406	5.308	5.003

Table 3: Estimates of the demand elasticity of markups, controlling for location-time effects

Dep. var.: log Price	(1) OLS	(2) TSLS	(3) TSLS	(4) TSLS
log Quantity	-0.047***	-0.055***	-0.177	-0.316*
	(0.001)	(0.003)	(0.114)	(0.172)
Observations	4958638	2029730	4955937	4924024
Products	1215	500	1215	1193
Product time	YES	YES	YES	YES
Product county	YES	YES	YES	YES
County time	YES	YES	YES	YES
Instrument	NO	Product	EXP	INC
1st stage F-stat		10.387	11.816	4.409

Notes: Standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SE are two-way clustered at the level of counties and products for specification (1), (3) and (4). For specification (2), the number of instruments would exceed the number of clusters. The SE for this specification are heteroskedasticity-robust. 1st stage F-statistic denotes the first stage Kleibergen-Paap rk Wald F statistic. Instruments are either product dummies, expensiveness or consumer income measures described in section 4. Specification (2) is estimated using a random draw of 500 products, while all other specifications include all products that satisfy sample selection criteria.

5.3 Robustness

First difference estimation In Table .A.4, I present estimates of the demand elasticity of markups using a first difference formulation of specification 2. Note that I difference all variables except for unemployment. Under nominal frictions or habit formation, both prices and quantities may be autocorrelated. Substantial differences between the baseline and first difference specifications would indicate that this is a problem for my estimation. The OLS and TSLS estimates using product-level dummies are very close to my baseline results. For the just-identified TSLS estimate using expensiveness as instrument, the coefficients are negative, significant, and very close to the baseline estimate of -0.5. For the just-identified TSLS estimate using consumer income as instrument, the estimate is positive, but insignificant and very imprecisely estimated. This is due to the fact that the consumer income instrument is even weaker in the first difference specification.

Allowing for segmented wholesale markets In Table .A.5, I present estimates that control for product-census division-time fixed effects. This specification relaxes the assumption of national wholesale prices to the assumption of census division specific wholesale prices. It thus includes census-division-product-time fixed effects instead of the product-time fixed effects of the baseline specification. This increases the number of estimated fixed-effects considerably. All point estimates are close to their values in the baseline specification. However, the instruments are significantly weaker, the TSLS standard errors are much larger, and the estimates for both just-identified TSLS specifications appear insignificant.

Allowing for category-specific production functions In Table .A.6, I present estimates that relax the assumption of an identical retail technology for all products to the assumption of identical technology for all products within one category (categories are frozen pizza, yogurt, shampoo, etc.). This requires me to estimate separate coefficients on w_{st} and y_{st} for each category. The causal effect of demand on markups is then identified through variation in $z_i y_{st}$ within a category of products. The table omits the category specific coefficients on w_{st} and y_{st} and only shows the coefficient of interest on q_{st}. Except for the TSLS estimate using consumer income as instrument, the estimates are very close to the baseline estimates, consistently negative and significantly different from zero.

6 Conclusion

In this paper, I estimate the response of markups to cyclical demand shocks. I show that positive product level demand shocks cause a decrease in markups. The effect is sizable, and my preferred estimate of the demand elasticity of markups is -0.5. This suggests that a 10% positive demand shocks reduces markups by 5 percentage points. The point estimates are imprecise, but consistently negative and significant in most specifications. My results support theories of counter-cyclical markups in which the price elasticity of demand varies with product-specific output, such as Ravn et al. (2006, 2008), and suggest that small demand shocks may be be considerably amplified by the counter-cyclical behavior of markups.

References

Banks, James, Richard Blundell, and Arthur Lewbel, "Quadratic Engel Curves and Consumer Demand," *The Review of Economics and Statistics*, 1997, *79* (4).

Bils, Mark, "The Cyclical Behavior of Marginal Cost and Price," *American Economic Review*, 1987, *77* (5).

_ **and James A. Kahn**, "What Inventory Behavior Tells Us about Business Cycles," *American Economic Review*, 2000, *90* (3).

_ **and Pete Klenow**, "Quantifying Quality Growth," *American Economic Review*, 2001, *4* (91).

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela, "The IRI Marketing Data Set," *Marketing Science*, July 2008, *27* (4).

Chevalier, Judith A. and David S. Scharfstein, "Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence," *The American Economic Review*, 1996, *86* (4).

_ **, Anil K. Kashyap, and Peter E. Rossi**, "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 2003, *93* (1).

Coibon, Oliver, Yuri Gorodnichenko, and Gee Hee Hong, "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications," *American Economic Review*, 2015, (105).

Gagnon, Etienne and David Lopez-Salido, "Small Price Responses to Large Demand Shocks," *Federal Reserve Board Finance and Economics Discussion Series*, 2014.

Gali, Jordi, Mark Gertler, and J. David Lopez-Salido, "Markups, Gaps, and the Welfare Costs of Business Fluctuations," *Review of Economics and Statistics*, 2007, *89* (1).

Gilchrist, Simon, Raphael Schoenle, Jae W Sim, and Egon Zakrajsek, "Inflation Dynamics During the Financial Crisis," *Society for Economic Dynamics Meeting Papers*, 2014.

Hall, Robert E, "What the Cyclical Response of Advertising Reveals about Markups and other Macroeconomic Wedges," *Working Paper*, April 2014.

Lewbel, Arthur and Krishna Pendakur, "Tricks with Hicks: The EASI Demand System," *American Economic Review*, 2009, *99* (3).

Nakamura, Alice O., Emi Nakamura, and Leonard I. Nakamura, "Price Dynamics, Retail Chains and Inflation Measurement," *Journal of Econometrics*, 2011, *161* (1).

Ramey, Valerie A., "Nonconvex Costs and the Behavior of Inventories," *Journal of Political Economy*, 1991.

Ravn, Morten O., Stephanie Schmitt-Grohe, and Martin Uribe, "Deep Habits," *Review of Economic Studies*, 2006, *1* (73).

_ , _ , and _ , "The Macroeconomics of Subsistence Points," *Macroeconomic Dynamics*, 2008, (12).

Rotemberg, Julio J. and Michael Woodford, "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity," *Journal of Political Economy*, 1992, *100* (6).

_ and _ , "The Cyclical Behavior of Prices and Cost," in John Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1 1999.

Staiger, Douglas and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 1997, *65* (3).

Stock, James H., Jonathan H. Wright, and Motohiro Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 2002, *20* (4).

Appendix

.A.1 Derivation of cost function

In this section I derive the cost function of stores operating as described in section 3. For simplicity I suppress the store and time subscripts. The production function for retail output of product i is given by $Q_i = L_i^{\alpha} X_i^{\beta} K_i^{1-\alpha-\beta}$. Where, L_i are the units of labor, X_i units of merchandise and K_i the share of capacity devoted to provision of product i. Factor prices are W and R_i. Total capacity K is fixed and normalized to 1, so that $\sum_i K_i = 1$. The cost minimization problem of a store providing $\{Q_i\}_{i=0}^{I}$ units of products $1, 2 \dots I$ can be written as follows:

$$\mathcal{L} = W \sum_{i=0}^{I} L_i + \sum_{i=0}^{I} X_i R_i + \sum_{i=0}^{I} \lambda_i (Q_i - L_i^{\alpha} X_i^{\beta} K_i^{1-\alpha-\beta}) + \phi(1 - \sum_{i=0}^{I} K_i)$$

The first order conditions are given by:

$$\begin{aligned}
W &= \lambda_i \alpha L_i^{\alpha-1} X_i^{\beta} K_i^{1-\alpha-\beta} && \forall i \in I \\
R_i &= \lambda_i \beta L_i^{\alpha} X_i^{\beta-1} K_i^{1-\alpha-\beta} && \forall i \in I \\
\phi &= -\lambda_i (1-\alpha-\beta) L_i^{\alpha} X_i^{\beta} K_i^{-\alpha-\beta} && \forall i \in I
\end{aligned}$$

It follows from combining the FOC, that:

$$\begin{aligned}
\frac{W}{R_i} &= \frac{\alpha}{\beta} \frac{X_i}{L_i} && \forall i \in I \\
K_i = \frac{L_i}{L} &= \frac{X_i R_i}{\sum X_i R_i} && \forall i \in I
\end{aligned}$$

We can now derive the product-specific factor demands:

$$\begin{aligned}
X_i &= Q_i \left(\frac{W}{R_i} \frac{\beta}{\alpha} \right)^{\alpha} R_i^{-(1-\alpha-\beta)} (X_i Z_i)^{1-\alpha-\beta} && \forall i \in I \\
L_i &= Q_i \left(\frac{W}{R_i} \frac{\beta}{\alpha} \right)^{-\beta} L^{1-\alpha-\beta} && \forall i \in I
\end{aligned}$$

The cost function follows from plugging factor demands into $C = W \sum_{i=0}^{I} L_i + \sum_{i=0}^{I} R_i X_i$:

$$C(\{Q_i\}, \{R_i\}, W) = \Omega W^{\frac{\alpha}{\alpha+\beta}} \left(\sum_{i=0}^{I} Q_i R_i^{\beta} \right)^{\frac{1}{\alpha+\beta}}$$

.A.2 Additional results

Table .A.4: First difference estimates of the demand elasticity of markups

	(1) OLS	(2) IVbase	(3) IVexpensiveness	(4) IVpcincome
Δ Quantity	-0.047***	-0.064***	-0.541***	0.374
	(0.001)	(0.011)	(0.155)	(0.329)
Unemployment	-0.000***	-0.000***	-0.002**	0.001
	(0.000)	(0.000)	(0.001)	(0.001)
Δ Wage	0.009***	0.010***	0.028	-0.007
	(0.003)	(0.004)	(0.022)	(0.018)
Observations	4524276	1852500	4524276	4496421
Products	1215	500	1215	1193
Product time	YES	YES	YES	YES
Product county	YES	YES	YES	YES
County time	NO	NO	NO	NO
Instrument	NO	Product	EXP	INC
widstat			12.179	2.064

Notes: Standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SE are two-way clustered at the level of counties and products for specification (1), (3) and (4). For specification (2), the number of instruments would exceed the number of clusters. The SE for this specification are heteroskedasticity-robust. 1st stage F-statistic denotes the first stage Kleibergen-Paap rk Wald F statistic. Instruments are either product dummies, expensiveness or consumer income measures described in section 4. Specification (2) is estimated using a random draw of 500 products, while all other specifications include all products that satisfy sample selection criteria.

Table .A.5: Estimates of the demand elasticity of markups, controlling for census division-product-time effects

	(1)	(2)	(3)	(4)
Dep. var.: log Price	OLS	TSLS	TSLS	TSLS
log Quantity	-0.040***	-0.046***	-0.520	-0.335
	(0.001)	(0.004)	(0.342)	(0.236)
Unemployment	-0.001	-0.001***	-0.007	-0.004
	(0.001)	(0.000)	(0.005)	(0.003)
log Wage	0.015**	0.016***	0.052	0.038
	(0.007)	(0.001)	(0.037)	(0.026)
Observations	4778601	1956544	4778601	4749181
Products	1215	500	1215	1193
Product time	YES	YES	YES	YES
Product county	YES	YES	YES	YES
County time	NO	NO	NO	NO
Instrument	NO	Product	EXP	INC
1st stage F-stat		5.705	2.524	2.556

Table .A.6: Estimates of the demand elasticity of markups, allowing for category-specific technology

	(1)	(2)	(3)	(4)
Dep. var.: log Price	OLS	TSLS	TSLS	TSLS
log Quantity	-0.044***	-0.049***	-0.584**	0.170
	(0.001)	(0.003)	(0.280)	(0.168)
Observations	4783401	1958571	4783401	4753813
Products	1215	500	1215	1193
Product time	YES	YES	YES	YES
Product county	YES	YES	YES	YES
County time	NO	NO	NO	NO
Instrument	NO	Product	EXP	INC
1st stage F-stat		9.302	4.746	3.297

Notes: Standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SE are two-way clustered at the level of counties and products for specification (1), (3) and (4). For specification (2), the number of instruments would exceed the number of clusters. The SE for this specification are heteroskedasticity-robust. 1st stage F-statistic denotes the first stage Kleibergen-Paap rk Wald F statistic. Instruments are either product dummies, expensiveness or consumer income measures described in section 4. Specification (2) is estimated using a random draw of 500 products, while all other specifications include all products that satisfy sample selection criteria.