

Markups Across Space and Time

Eric Anderson*, Sergio Rebelo[†], and Arlene Wong[‡]

April 6, 2018

Abstract

In this paper, we provide direct evidence on the behavior markups in the retail sector across space and time. Markups are measured using gross margins. We consider three levels of aggregation: the retail sector as a whole, firm-level data, and product-level data. We find that: (1) markups are relatively stable over time and mildly procyclical; (2) there is large regional dispersion in markups; (3) there is a positive cross-sectional correlation between local income and local markups; and (4) differences in markups across regions are explained by differences in assortment, not by deviations from uniform pricing. We propose an endogenous assortment model that is consistent with these facts.

J.E.L. Classification: E30

Keywords: Gross margins, prices, marginal costs, business cycle.

*Northwestern University.

[†]Northwestern University, NBER, and CEPR.

[‡]Princeton University, and NBER.

1 Introduction

Are markups procyclical, acyclical or countercyclical? The answer to this question is an important input into the evaluation of macro models. Most empirical studies of the cyclical properties of markups use structural approaches that rely on assumptions about production functions and market structure.¹ The literature is divided in its conclusions, in part because different studies resort to different assumptions.

In this paper, we provide direct empirical evidence on the cyclical properties of markups based on gross margins for the retail industry. We focus on the retail sector because its predominant variable cost, the cost of goods sold, can be used as a proxy for marginal cost.² Moreover, estimates of the frequency of price changes and other statistics computed using retail prices have often been used to evaluate the importance of nominal rigidities and to calibrate macroeconomic models.³

We study markups at three levels of aggregation: for the retail sector as a whole, at the firm level, and at the product level. The product-level analysis is based on scanner data from a large retailer. This data set has three important advantages. First, it includes the price of every transaction, instead of the average price across transactions. Second, it contains the replacement cost of every item, which is a good proxy for marginal cost. Third, the data pertains to stores located in different regions, which allows us to study the regional distribution of markups.

Our main empirical finding is that gross margins are relatively stable over time and mildly procyclical. In contrast, sales and net operating margins are strongly procyclical and quite volatile. These results are consistent across all three levels of aggregation, for the aggregate

¹For example, [Bils \(1987\)](#), [Hall \(1988\)](#), [Galeotti and Schiantarelli \(1998\)](#), [Rotemberg and Woodford \(1999\)](#), [Bils and Kahn \(2000\)](#), [Kryvtsov and Midrigan \(2012\)](#), [Nekarda and Ramey \(2013\)](#), [Karabarbounis \(2014\)](#), [Kim \(2015\)](#), and [Bils, Klenow and Malin \(2017\)](#) infer the cyclical properties of mark-ups from the cyclical properties of the cost shares of labor and other inputs. [Hall \(2014\)](#) bases his analysis on the cyclical properties of advertising expenses. [Collard-Wexler and Loecker \(2015\)](#) infer the cyclical properties of markups using their estimates of the production function. See [Nekarda and Ramey \(2013\)](#) for a detailed summary of studies that find countercyclical, procyclical and acyclical markups.

²See [Eichenbaum, Jaimovich and Rebelo \(2011\)](#) for a discussion of the conditions under which the cost of goods sold is a good proxy for the marginal cost. [De Loecker and Eeckhout \(2017\)](#) use gross margins to estimate long-term trends in markups.

³[Bils and Klenow \(2004\)](#) is a seminal paper in the use of retail prices from the consumer price index to study the importance of nominal rigidities. See [Klenow and Malin \(2011\)](#) for a review of the micro price literature.

retail sector, at the firm level, and at the product level. Our product-level evidence suggests that the marginal replacement cost of good sold is relatively stable.

We also find that the response of gross margins to monetary policy shocks and oil shocks is not statistically different from zero. In contrast, the response of net operating margins to these shocks is negative and statistically significant.

The relative stability of gross markups over time contrasts sharply with the large regional dispersion in gross markups implied by our scanner data set. This regional dispersion is driven by differences in prices not by differences in marginal costs. We find that regions with higher incomes and more expensive houses tend to pay higher markups. These higher markups reflect differences in assortment rather than regional differences in markups charged for the same item. In other words, high-income regions pay higher markups on an assortment of goods that is different from the assortment offered and sold in low-income regions. Items sold in both high- and low-income regions generally have uniform prices. Our regional evidence suggests that permanent shocks might result in permanent changes in assortment and markups.

We consider a broad and a narrow interpretation of the implications of our empirical findings for the evaluation of macro models. The broad interpretation relies on the strong assumption that the cyclical properties of markups that we document for the retail sector can be generalized to the rest of the economy. Under this interpretation, our evidence is inconsistent with models with countercyclical markups. This class of models includes the standard New-Keynesian models with sticky prices and procyclical marginal costs (see e.g. Woodford (2003)), Ravn, Schmitt-Grohé, and Uribe (2008)'s deep-habit model, and Jaimovich and Floetotto (2008)'s entry and exit model. Our evidence favors models that generate acyclical or weakly procyclical markups. This class of models includes economies with both sticky prices and wages (see Erceg, Henderson, and Levin (2000), Christiano, Eichenbaum and Evans (2005) and Christiano, Eichenbaum and Trabandt (2015)).

The narrow interpretation of our findings is that they apply only to the retail sector. Under this interpretation, our evidence is consistent with models with nominal rigidities (sticky prices or wages) at the manufacturing level that keep manufacturing prices relatively stable, combined with either flexible (as in Coibion, Gorodnichenko and Hong (2015)) or sticky prices (as in Nakamura and Steinsson (2010) and Pasten, Schoenle and Weber (2016))

at the retail level. Our evidence is inconsistent with models such as Golosov and Lucas (2007) and Midrigan (2011). Golosov and Lucas (2007) model vertically integrated retailers subject to menu costs. In Midrigan’s (2011) model, retailers have sticky prices due to menu costs but there are no nominal rigidities in manufacturing. Both of these models imply that retail markups are countercyclical and marginal costs of replacing goods sold are volatile.

Existing macro models are generally inconsistent with the regional correlation between markups and income present in our data. The trade models proposed by Fajgelbaum, Grossman and Helpman (2011) and Bertolotti and Etro (2017), which feature non-homothetic preferences, are consistent with this regional correlation.

We present a simple variant of the Dixit-Stiglitz model that draws on insights by Fajgelbaum, Grossman and Helpman (2011) and is consistent with both our time-series and regional evidence. Our model has the following properties: (1) gross margins are relatively stable over time and mildly procyclical; (2) there is large regional dispersion in gross margins; (3) there is a positive cross-sectional correlation between local income and local gross margins; and (4) differences in gross margins across regions are explained by differences in assortment, not by deviations from uniform pricing.

In sum, we provide direct empirical evidence on the behavior of markups, as well as a theory that is consistent with our findings. This paper is organized as follows. Section 2 describes the data we use. Section 3 contains our empirical results. Section 4 discusses the implications of our findings for business cycle and trade models. This section also presents an endogenous assortment model consistent with our findings. Section 5 concludes.

2 Data

Our analysis focuses on the retail sector, which accounts for roughly 10 percent of aggregate employment. We use two data sets. The first, obtained from Compustat, includes quarterly panel data on sales, costs of goods sold, selling and administrative expenses, and net profits for retail firms for the period from 1979 to 2014 . We have 1,735 retail firms in our sample. In the Appendix, we show that the sales growth rates from the Compustat data for the retail sector track closely the sales growth rates obtained from the U.S. Census Retail survey data.

Using these data, we construct two key margins for each firm f in quarter t :

$$(\text{Gross margin})_{ft} = \frac{\text{Sales}_{ft} - (\text{Cost of goods sold})_{ft}}{\text{Sales}_{ft}} \quad (1)$$

$$\begin{aligned} (\text{Net operating profit margin})_{ft} &= \frac{\text{Sales}_{ft} - (\text{Cost of goods sold})_{ft} - (\text{Other expenses})_{ft}}{\text{Sales}_{ft}} \\ &= (\text{Gross margin})_{ft} - \frac{(\text{Other expenses})_{ft}}{\text{Sales}_{ft}} \end{aligned} \quad (2)$$

Other expenses include overhead expenses, rent, labor costs, and capital and property depreciation. For retail firms, these expenses are predominately fixed or quasi-fixed costs.

Our second data source is a scanner data set from a large retailer that operates more than 100 stores in different U.S. states. This retailer sells products in the grocery, health and beauty, and general merchandise categories. We have weekly observations on quantities sold, retail and wholesale prices for each item in each of the retailer’s stores. An item is a good, defined by its stock keeping unit code (SKU) in a particular store. In total, we have roughly 3.6 million SKU-store pairs across 79 product categories. Our sample period begins in the first quarter of 2006 and ends in the third quarter of 2009, so it includes the recession that started in the 4th quarter of 2007 and ended in the second quarter of 2009.

Using the scanner data, we construct the gross margin for each item, i , at store s , in county k , at time t :

$$(\text{Gross margin})_{iskt} = \frac{\text{Price}_{iskt} - (\text{Replacement cost})_{iskt}}{\text{Price}_{iskt}}. \quad (3)$$

Since the real GDP data we use to measure economic activity is available quarterly, we construct gross margins at a quarterly frequency by expenditure-weighting weekly gross margins.

Our data set has two key features that distinguish it from a number of other scanner data sets.⁴ First, it contains the price of every transaction, instead of the average price across transactions. Second, the cost data measures the replacement cost, which is a good proxy for the marginal cost. Moreover, the marginal cost is available at the store level, rather than

⁴Data from this retailer have been used in other studies, including Anderson, Malin, Nakamura, Simester, and Steinsson (2016), Anderson, Jaimovich and Simester (2015) and McShane et al. (2016).

as a national average. This property allows us to compute the markup above marginal cost for each item and store at each point in time.

We also use data on the unemployment rate, real GDP growth, and estimates of monetary policy and oil price shocks. The monetary policy shocks are identified from high-frequency Federal Funds futures data.⁵ Oil price shocks are identified using the approach proposed by Ramey and Vine (2010). We provide additional details on the process used to estimate these shocks in the Appendix.

3 Business cycle properties

This section documents the cyclical properties of gross margins, operating margins, sales, and cost of goods sold. We discuss the comovement and volatility of these series for the aggregate retail sector, at the firm level, and at the product level.

3.1 Aggregate retail sector evidence

We construct aggregate measures of our variables for the retail sector using aggregate sales and aggregate costs. Table 1 summarizes the elasticity of different variables with respect to real GDP. This elasticity is estimated by regressing the year-on-year logarithmic difference of each variable on the year-on-year logarithmic difference of real GDP.

We see that gross margins are roughly acyclical or mildly procyclical. In contrast, sales and cost of goods sold are highly procyclical. These properties suggest that firms are not changing markups in response to business cycle fluctuations. Rather, the business cycle affects primarily their quantities sold and any cost increases from their suppliers, which is why sales and cost of goods sold are highly procyclical.

Table 2 shows that gross margins are relatively stable when compared to other variables. At a quarterly frequency, operating profit margins are 3.4 times more volatile than gross margins, while sales and costs are roughly 2.6 times more volatile than gross margins. The high volatility of operating profit margins compared to the volatility of gross margins suggests

⁵See Kuttner (2001) and Gurkaynak, Sack, and Swanson (2005) for details on the construction of these shocks.

that fixed costs might be an important driver of profitability. Figure 1 which depicts the log-differences from the prior year of gross margins and operating margins, illustrates the differences in volatility between these two variables.

3.2 Firm-level evidence

To study the cyclical properties of the firm-level variables, we regress each variable on the year-on-year log-difference in real GDP using firm fixed effects. Our firm fixed effects takes out any permanent differences across firms, including differences in the degree of vertical integration between the retail and manufacturing operations.

Table 3 reports the resulting elasticities. The elasticity of the gross margin is small and statistically insignificant, while the elasticities of operating profits, sales and cost of goods sold are positive and statistically significant. Consistent with the aggregate evidence, the firm-level evidence suggests that business cycles primarily affect the firms' quantities sold and any cost increases from suppliers, rather than their gross margins.

To study the volatility of a given variables at the firm level, we estimate the standard deviation of this variable for each firm and then compute the average of this statistic across firms. We report our results in Table 4. The operating profit margin is the most volatile variable in our sample while the gross margins is the least volatile.

Finally, we study the conditional response of the gross margin and the operating profit margin to high-frequency monetary-policy shocks and oil-price shocks.⁶ We estimate this response by running the following regression separately for the gross margin and the net operating profit margin:

$$\Delta \ln m_{it} = \beta_0 + \sum_k \beta_k \epsilon_{t-k} + \lambda_{q(t)} + \lambda_r + \eta_{it},$$

where $\Delta \ln m_{it}$ is the year-on-year log-difference in the margin of firm i at time t . The variable ϵ_{t-k} is the aggregate shock at time $t - k$. The variables $\lambda_{q(t)}$, λ_r , and λ_i are fixed effects for the calendar quarter, recession, and firm.

⁶Our scanner data does not contain enough time periods to allow us to estimate the conditional response of the gross margin to shocks.

Figure 2 depicts the implied impulse response functions. We see that the response of the gross margin is statistically insignificant for both monetary and oil-price shocks. In contrast, net operating profit margins fall in a statistically significant manner in response to both shocks.

These results are at odds with the properties of simple New-Keynesian models. These models generally predict that gross margins rise in response to monetary shocks and fall in response to oil-price shocks. Monetary shocks are contractionary, so they produce a fall in marginal costs. Since prices are relatively stable, the gross margin rises. Oil-price shocks are also contractionary, but they produce a rise in marginal costs and a fall in the gross margin. In our data, the gross margin does not respond to either monetary or oil-price shocks.

3.3 Product-level evidence

There are two potential sources of measurement error associated with our aggregate data for the retail sector. First, gross margins are constructed using average costs instead of marginal costs. Second, changes in inventories affect the cost of goods sold and can potentially affect the cyclical properties of our empirical measure of the gross margin.⁷ We now report results that are free of these two sources of measurement error.

Our analysis is based on a scanner data set from a large retailer which includes transaction prices and replacement costs at the SKU level. Using this information, we compute gross margins for every product in every store. We aggregate the weekly observations to construct quarterly data.

We use our product-level data to show that the gross margins based on the cost of goods sold used in the previous subsections are a good proxy for gross margins based on the marginal replacement cost. We find that the correlation between the two measures of gross margins is 96 percent.

Figure 3 shows how the retailer reacted to the onset of the 2009 recession. This figure plots the regional distribution in the level of markup and in the year-on-year log difference in sales and number of unique items for the periods 2006-07 and 2008-09. For confidentiality

⁷In Appendix A2, we present a version of our analysis where we adjust the cost of goods sold for changes in inventories. We still find that the elasticity of gross margins with respect to GDP is statistically insignificant.

reasons, we do not report the level of the average gross margin. In constructing Figure 3, we normalize the gross margins by subtracting the average gross margin for the period 2006-07 from the gross margins for 2006-07 and 2008-09. As a result, the normalized average gross margin for the period 2006-07 is zero.

We see that the regional distribution of the level of gross margins remained relatively stable with a small shift to the left. In contrast, the distribution of year-on-year log difference in sales is more skewed in the Great Recession than in the 2006-07 period.⁸ The distribution of the number of unique items in each store shifted to the left. In other words, the retailer reacted to the drop in sales associated with the recession by reducing the number of unique items in each store while leaving gross margins roughly stable.

Table 5 reports the average, median, 10th and 90th percentiles of the distribution of the three variables in Figure 3 for the expansion and recession periods. The gross-margin moments are similar across the two periods.⁹ In contrast, the sales and number of item moments are all lower in the recession.

To go beyond these unconditional moments we now compute the elasticity of the variables of interest with respect to the local rate of unemployment and local real house prices.¹⁰ We estimate the following regression:

$$\Delta \log \text{margins}_m = \beta_0 + \beta_1 \Delta \log(Z_t) + \gamma X_m + \varepsilon_m,$$

where m denotes the region and the variables are log-differences between the period 2005-06 and 2007-08. We consider two possible alternative explanatory variables, Z_t : the local unemployment rate, and house prices instrumented with the housing supply elasticity from Saiz (2010).¹¹ The vector X_m represents a set of controls including local area income, racial composition, median age, manufacturing industry share of employment, and share of college-educated workers.

⁸This result is consistent with the findings of Bloom, Guvenen and Salgao (2015). These authors show that sales growth becomes skewed during recessions in several countries.

⁹For confidentiality reasons, we do not report the average gross margin, only the difference in the average gross margin across the expansion and recession period.

¹⁰We thank Emi Nakamura for sharing with us data on unemployment for the regions included in our scanner data.

¹¹This instrument uses information on the geography of a metropolitan area to measure the ease with which new housing can be constructed. The index assigns a high elasticity to areas with a flat topology and without many water bodies, such as lakes and oceans. In areas with low elasticity, it is more difficult to expand the housing supply in response to a demand shock, and house prices should therefore increase by more.

Table 6 reports our results. The elasticity of the gross margin is statistically insignificant with respect to unemployment and it is positive and statistically significant with respect to local house prices. The price and replacement cost elasticity are also statistically insignificant at a 5 percent confidence level. The elasticity of sales is statistically significant for both the unemployment rate and local house prices, indicating that sales rise during periods when the local economy booms. Finally, the number of unique items carried in the store is procyclical; its elasticity is statistically significant at a 1 percent confidence level.

Table 7 shows the standard deviation of year-on-year logarithmic changes in different variables. We see that markups, prices and cost of goods sold are relatively stable. In contrast, sales and the number of unique items in store's assortment are quite volatile.

3.4 Summary

In this section we study the cyclical properties of the retail sector at three levels of aggregation: for the retail sector as a whole, for individual firms in the retail sector and at the level of the product, using data for one large retailer.

Our main findings for aggregate and firm-level data are as follows. Gross retail margins are stable over the business cycle and mildly procyclical. In contrast, sales, cost of goods sold, and net operating profits are highly procyclical. The high volatility of net operating costs is suggestive of the presence of large fixed costs. Operating profit margins are much more volatile than gross margins which suggests the presence of fixed costs.

The evidence for our large retailer indicates that the firm reacted to the 2009 recession primarily by reducing the number of stores and the number of unique items. Presumably, these actions result from reduction in the number of stores reflects the desire to reduce fixed costs. Gross margins remained relatively stable, falling slightly.

4 Cross-sectional properties

We can use our scanner data to study the cross-sectional distribution of the level of gross margins across regions. In the previous section, we use Figure 3 to show that the regional distribution of gross margins is relatively similar in the Great Recession and in the expansion

that preceded it. The same figure shows that there is a large regional dispersion in the markups charged by our large retailer in both the expansion and the recession period.

We can decompose the overall variance in the gross margins into a time-series and a regional component. We denote by v_{mt} the gross margin of region m at time t , computed as a sales-weighted average of all items in stores located in this region. The variance of v_{mt} can be written as:

$$\begin{aligned} \text{var}(v_{mt}) &= \frac{1}{TM-1} \sum_t \sum_m (v_{mt} - v)^2 \\ &= \frac{1}{TM-1} \sum_t \sum_m (v_{mt} - v_t + v_t - v)^2 \\ &\approx \frac{1}{T} \sum_t \underbrace{\frac{\sum_m (v_{mt} - v_t)^2}{M-1}}_{\text{var}_t(v_m)} + \underbrace{\frac{\sum_t \sum_m (v_t - v)^2}{TM-1}}_{\text{var}(v_t)} + 2\text{cov}(v_{mt} - v_t, v_t - v), \end{aligned}$$

where T is the total number of time periods and M is the total number of regions. The variable v_t is the average gross margin across all regions at time t , computed as a sales-weighted average of all items in all stores. The variable v is the average of v_t across time. The variables $\frac{1}{T} \sum_t \text{var}_t(v_m)$ and $\text{var}(v_t)$ represent the average regional variance and the time-series variance in gross margins, respectively. The variable $\text{cov}(v_{mt} - v_t, v_t - v)$ is the covariance between the time-series and the regional component.

The regional variance in markups, $\frac{1}{T} \sum_t \text{var}_t(v_m)$, is 0.26 while the time-series variation, $\text{var}(v_t)$, is 0.04. The covariance term, $\text{cov}(v_{mt} - v_t, v_t - v)$, is -0.0015 . This decomposition suggests that most of the variation in markups comes from the cross section, not from the time series.

To study the source of regional variation in markups, we start with the following equation for the variance of markups across different markets conditional on period t , $\text{var}_t(v_m)$:

$$\text{var}_t(v_m) = \text{var}_t \left(\sum_j v_{jm} w_{jm} \right). \quad (4)$$

Here, v_{jm} is the markup of product j in market m and w_{jm} is the sales of product j in market m as a fraction of total sales in market m .

Expanding the terms on the right-hand side of equation (4), we obtain:

$$\begin{aligned}
var_t(v_m) = & \underbrace{var_t \left[\sum_j (v_{jm} - \bar{v}_j) \bar{w}_j \right]}_{\text{differences in gross margins for the same item}} \\
& + \underbrace{var_t \left[\sum_j (w_{jm} - \bar{w}_j) \bar{v}_j \right]}_{\text{differences in assortment composition}} \\
& + \underbrace{var_t \left[\sum_j (v_{jm} - \bar{v}_j)(w_{jm} - \bar{w}_j) \right]}_{\text{interaction term}} + \text{covariance terms.}
\end{aligned}$$

The first term on the right-hand side of this equation measures the importance of differences in gross margins for the same item. This term is zero when there is uniform pricing, i.e. prices for the same product are identical in different regions. The second term measures the importance of differences in assortment holding fixed the gross margin across regions. This term is zero when all regions have the same assortment composition. The third term measures the importance of the interaction between differences in assortment and differences in gross margins.

Table 8 reports two versions of this decomposition. For both versions we report average estimates across time. The first column shows the results we obtain when we restrict the sample to items that are sold in every market. The second column reports results obtained using all items, including items that are sold only in a subset of the regions. In both cases, we find that the predominant driver of regional differences in gross margins are differences in assortment composition across regions. In contrast, regional differences in the markups of the same items account for very little of the regional variation in gross margins. In other words, when the same item is available in different regions our retailer uses roughly uniform pricing.

Table 9 shows that gross margins are positively correlated with measures of income or wealth. These measures include the logarithm of household income, the logarithm of median house value, the share of income received by the top 1 percent. In contrast, gross margins are uncorrelated with a measure of competition (the Herfindahl index) and a proxy for higher transportation costs (a dummy variable that takes the value one for counties classified by the census as rural).

Our results help reconcile the findings of Stroebel and Vavra (2016) with those of Della Vigna and Gentzkow (2017). Stroebel and Vavra (2016) find that prices are correlated with wealth, which is proxied by house-price shocks. They use two proxies for marginal cost (wages and wholesale prices) and find that these do not vary across regions. They interpret the positive correlation of prices with wealth as reflecting higher markups, reflecting the fact that wealthier households have lower price elasticities of demand. Della Vigna and Gentzkow (2017) find evidence of uniform pricing and that individual prices do not respond to local shocks.

We find that there is indeed a positive cross-sectional correlation between local income and local gross margins. But these differences in gross margins across regions are explained by differences in assortment, not by deviations from uniform pricing. Our findings are consistent with recent work by Neiman and Vavra (2018) who show that households concentrate their spending on different goods. We add to their results by providing direct evidence on differences in markups and assortment across regions.

5 Macroeconomic and trade models

In this section, we evaluate several business cycle and trade models in light of our evidence. We then present an endogenous assortment model that is broadly consistent with both our time-series and cross-sectional evidence.

5.1 Business cycle models

In the introduction, we discuss a broad and narrow interpretation of our findings. Under a broad interpretation that our evidence is applicable to the economy as a whole, our results favor models that generate acyclical or weakly procyclical markups. Prominent examples of this class of models are economies with both prices and wage rigidities (e.g. Erceg, Henderson, and Levin (2000), Christiano, Eichenbaum and Evans (2005) and Christiano, Eichenbaum and Trabandt (2015)). Under a narrow interpretation that our findings only apply to the retail sector, our results favor models with nominal rigidities (sticky prices or wages) at the manufacturing level that keep manufacturing prices relatively stable, combined

with either flexible (as in Coibion, Gorodnichenko and Hong (2015)) or sticky prices (as in Nakamura and Steinsson (2010) and Pasten, Schoenle and Weber (2016)) at the retail level.

However, none of the models mentioned above are consistent with our finding that markups are correlated with income in the cross section.

5.2 Trade models

Trade models with non-homothetic preferences generate a positive correlation between markups and income. Bertolotti and Etro (2017) consider a version of the Dixit-Stiglitz model of monopolistic competition with a non-homothetic aggregator. Fajgelbaum, Grossman and Helpman (2011) propose a model with non-homothetic preferences in which households consume an homogeneous good and a single unit of a differentiated good. Households choose the quantity of the homogeneous good and the quality of the differentiated good. We discuss the properties of these two models in turn. Both models are static so income and consumption expenditures coincide.

5.2.1 The Bertolotti and Etro model

Bertolotti and Etro (2017) write the household's indirect utility function as:

$$\int_0^n \mu(p_i/Y) di,$$

where p_i denotes the price of differentiated good i and Y represents income. The authors show that when $\mu(\cdot)$ takes an exponential form,

$$\mu(p_i/Y) = \exp[-\tau (p_i/Y)],$$

the markup of price over marginal cost (c) is given by:

$$\frac{p_i}{c} = 1 + \frac{Y}{\tau c}.$$

When $\mu(\cdot)$ takes an addilog form,

$$\mu(p_i/Y) = [a - (p_i/Y)]^{1+\gamma},$$

the markup of price over marginal cost (c) is given by:

$$\frac{p_i}{c} = \frac{\gamma + a(Y/c)}{1 + \gamma}.$$

Consistent with our time-series evidence, as long as the cyclicality of income and marginal costs is similar, markups are roughly acyclical. The model is also consistent with our cross-sectional evidence. Suppose that marginal costs are similar across regions but there is dispersion in income levels. Then, higher income regions pay higher markups.

However, this model is inconsistent with the nature of the regional variation in markups present in our data. Our evidence suggests that markups vary with income or wealth because richer regions buy an assortment of goods that is different from poor regions. In contrast, the Bertolotti and Etro (2017) model implies that regions with different levels of income have different markups for the *same* item.

5.2.2 The Fajgelbaum, Grossman and Helpman model

The model proposed by Fajgelbaum, Grossman and Helpman (2011) is fully consistent with our cross-sectional evidence under the assumption that there is less substitutability between brands of higher quality than between brands of lower quality. Under this assumption, the model implies that regions with higher income pay higher markups but consume higher quality items. So variations in markups are driven by differences in assortment, just like in our scanner data.

Unfortunately, the Fajgelbaum, Grossman and Helpman (2011) model is inconsistent with our time-series evidence. The markup over marginal cost (c_i) for an item of quality q_i and brand j is:

$$\frac{p_{ij}}{c_i} = 1 + \frac{\theta_i}{q_i c_i},$$

where θ_i is the dissimilarity parameter. This formula implies that, when marginal costs are procyclical, the model generates countercyclical markups for each item i .

A version of the Fajgelbaum, Grossman and Helpman with sticky wages might be consistent with both the time-series and cross-sectional evidence. But such a model would have complex borrowing and lending across agents that would greatly reduce its tractability. Instead of pursuing this route, we consider a version of the Dixit-Stiglitz model that

embodies a central insight from Fajgelbaum, Grossman and Helpman (2011): higher quality consumption bundles are made of less substitutable components.

5.3 An endogenous assortment model

Our economy is populated by a representative household who maximizes its lifetime utility given by:

$$U = E_0 \sum_{t=0}^{\infty} \left\{ \beta_t \left[\log(C_t^\alpha Z_t^{1-\alpha}) + \theta_t \log(1 - N_t) \right] \right\}. \quad (5)$$

The symbol E_0 denotes the expectation conditional on the information set available at time zero. The variables N_t and Z_t denote hours of labor and the consumption of an homogenous good, respectively. The variable θ_t represents a shock to the labor supply.

A consumption bundle, C_t , with quality q_t , is a composite of n_t differentiated goods combined according to a Dixit-Stiglitz aggregator:

$$C_t = q_t^\gamma \left[\int_0^{n_t} x_{iqt}^{1/v(q_t)} di \right]^{\nu(q_t)},$$

where x_{iqt} is the quantity consumed of variety i with quality q at time t . We assume that $v(q_t)$ is an increasing function of q_t . So, as in Fajgelbaum, Grossman and Helpman (2011), higher-quality consumption bundles are produced with more differentiated inputs.

For tractability, we consider the simple case in which $\nu(q_t)$ is a linear function, so v_t is equal to the quality of the inputs ($v_t = q_t$) and the consumption aggregator is given by:

$$C_t = v_t^\gamma \left[\int_0^n x_{ivt}^{1/v_t} di \right]^{\nu_t}.$$

We assume that $\gamma > 1$ which implies that, other things equal, households prefer higher quality baskets. We also assume that there's a minimum consumption size for each variety. For convenience, we normalize this minimum size to one:

$$x_{ivt} \geq 1.$$

We can solve the household's problem in two steps. The first step is to find the efficient consumption of varieties, minimizing total expenditure, for a given level of C_t , \bar{C}_t :

$$\min_{x_{ivt}, v_t} \int_0^{n_t} p_{ivt} x_{ivt} di,$$

subject to:

$$\bar{C}_t = v_t^\gamma \left[\int_0^n x_{ivt}^{1/v_t} di \right]^{\nu_t}.$$

Households choose the quality of the consumption bundle, q_t , and the amount consumed of each individual variety with quality q_t , x_{ivt} . The first-order conditions for this problem are:

$$\frac{x_{ivt}}{x_{jvt}} = \left[\frac{p_{ivt}}{p_{jvt}} \right]^{\nu_t/(1-\nu_t)}.$$

The elasticity of substitution between any two varieties is $-\nu_t/(1-\nu_t) \geq 0$. The case of $\nu_t = \infty$ corresponds to the Cobb-Douglas case. Finite values of ν_t are associated with a lower elasticity of substitution than Cobb-Douglas.

The optimal allocation of the differentiated consumption goods satisfies the condition,

$$p_{ivt} = \nu_t^{\gamma/v_t} P_t C_t^{(v_t-1)/v_t} x_{ivt}^{(1-\nu_t)/v_t}.$$

Here, P_t is the price index associated with the bundle C_t :

$$P_t = v_t^{-\gamma} \left[\int_0^n p_{ivt}^{1/(1-\nu_t)} di \right]^{1-\nu_t}. \quad (6)$$

The second step is to maximize lifetime utility subject to the household's budget constraint. The household's income, Y_t , is given by the sum of labor income and firm profits:

$$Y_t = w_t N_t + \int_0^{n_t} \pi_{it} di.$$

The household budget constraint is:

$$Y_t = \int_0^{n_t} x_{ivt} p_{ivt} di + Z_t.$$

We choose the homogeneous good as the numeraire, so its price is one. The first-order conditions for this problem are:

$$\begin{aligned} \frac{\theta}{1-N_t} &= (1-\alpha) \frac{w_t}{Z_t}, \\ P_t C_t &= \alpha Y_t, \\ Z_t &= (1-\alpha) Y_t. \end{aligned}$$

Production Each intermediate good of quality ν_t is produced with labor:

$$x_{ivt} = A_t(1 + g)^t N_{ivt},$$

where A_t is a stationary shock to productivity and g is the long-run growth rate of productivity.

The monopolist of variety i supplies the level of quality demanded by consumers. Its problem is to maximize profits given by:

$$\pi_{it} = p_{ivt}x_{ivt} - \frac{w_t}{A_t(1 + g)^t}x_{ivt} - \Psi, \quad (7)$$

where Ψ denotes a fixed cost that the firm must incur in every period of operation.

The optimal price is given by the usual markup equation:

$$p_{ivt} = \nu_t \frac{w_t}{A_t(1 + g)^t}.$$

Producers of the homogeneous good The homogeneous good is produced by competitive producers using labor and the following production function:

$$Z_t = (1 + g)^t N_{zt}.$$

We assume that there is a continuum of measure one of homogeneous-good producers. The problem of the representative producer is to maximize:

$$\pi_{zt} = Z_t \left[1 - \frac{w_t}{(1 + g)^t} \right].$$

Real income It is useful to define real income, \tilde{Y}_t , measured in terms of the consumption basket of differentiated and homogeneous goods purchased by the households:

$$\tilde{Y}_t = \frac{Y_t}{P_t^\alpha}. \quad (8)$$

Recall that α is the share of the bundle of differentiated goods in household expenditure.

Equilibrium In equilibrium, households maximize their utility, (5), taking the wage rate and prices as given. Monopolists maximize profits taking the wage rate, the aggregate consumption bundle, C_t , and the aggregate price of the bundle of consumption varieties, P_t , as given. Producers of the homogeneous good maximize profits, taking prices as given. The labor market clears:

$$N_{zt} + \int_0^{n_t} N_{ivt} di = N_t.$$

The market for differentiated goods and for the homogeneous good clear.

Using the household budget constraint, we can rewrite C_t in a symmetric equilibrium as:

$$C_t = \nu_t^{\gamma-1} n_t^{\nu-1} \alpha Y_t A_t.$$

Since $\gamma > 1$, the household's utility is monotonically increasing in ν_t . The value of x_{vt} is given by:

$$x_{vt} = \frac{\alpha A_t Y_t}{\nu_t n_t}. \quad (9)$$

Since utility is increasing in ν_t , the constraint $x_{vt} \geq 1$ is binding. Setting $x_{vt} = 1$ in equation (12), we obtain the optimal value of ν_t :

$$\nu_t = \frac{\alpha A_t Y_t}{n_t}. \quad (10)$$

The following proposition, proved in the Appendix, summarizes the properties of the equilibrium.

Proposition 1. *The equilibrium of this economy is described by the following equations:*

$$\begin{aligned} w_t &= (1+g)^t, \\ Y_t &= \frac{(1+g_t)^t}{1+\theta_t}, \\ n_t &= \frac{\alpha A_t (1+g_t)^t}{(1+\Psi A_t)(1+\theta_t)}, \\ x_{ivt} &= 1, \\ p_{ivt} &= \Psi + 1/A_t, \\ N_t &= \frac{1}{1+\theta_t}, \\ \nu_t &= 1 + \Psi A_t. \\ \tilde{Y}_t &= \frac{A_t^\alpha \left(\frac{\alpha}{\Psi+1/A_t} \frac{1}{1+\theta_t} \right)^{\alpha \Psi A_t}}{(1+\Psi A_t)^{(1-\gamma)\alpha}} \frac{1}{1+\theta_t} \left[(1+g)^{1+\alpha \Psi A_t} \right]^t \end{aligned}$$

Real income, \tilde{Y}_t , is an increasing function of A_t and a decreasing function of θ_t .

To study the model's steady state properties, suppose that A_t and θ_t are constant. The price of each differentiated good, hours worked and the markup are also constant. Real wages, household income measured in units of the homogeneous good, and the number of firms producing differentiated goods grow at a constant rate g .

Real income measured in terms of the consumption basket, \tilde{Y}_t , grows at a gross rate of $(1 + g)^{1 + \alpha\Psi A_t}$. The reason this gross rate is higher than $1 + g$ is as follows. Equation (6) shows that the price index for differentiated goods is proportional to n_t^{1-v} which, in equilibrium, equals $n_t^{-\Psi A_t}$. The number of firms grows at a gross rate $1 + g$, increasing variety and changing the effective price of the basket of differentiated goods at a gross rate $(1 + g)^{-\Psi A_t}$. Since differentiated goods have a weight of α in the overall consumer basket, growth in variety results in a fall in the basket price and a rise in real income of $(1 + g)^{\alpha\Psi A_t}$.

When the fixed cost ψ is zero, the equilibrium value of v is one. The differentiated goods are perfect substitutes and so the net markup is zero, price equals marginal cost. One possible interpretation of ψ is that it corresponds to the rental costs associated with operating in a given region. When ψ is high, there are fewer firms in equilibrium. Since it is optimal for households to consume only one unit of each differentiated good, households adjust to a lower number of firms by consuming higher-quality goods. This higher quality corresponds to a higher markup that allows firms to recoup the high value of the fixed costs.

Model implications To assess the model's regional implications, we compare regions that have different productivity levels and thus different levels of real income. Higher productivity regions have higher markups and a higher number of varieties. This implication is consistent with the finding we report on Section 4: gross margins and the number of varieties are positively correlated with income.

To assess the model's cyclical properties, we consider the effects of temporary shocks to productivity and labor supply. Consider first the effect of an increase in A_t . Households increase the quality of the varieties they consume and, as a result, the markup for differentiated goods increases.¹² Profits would rise if the number of firms stayed constant. In

¹²See Bils and Klenow (2001) for evidence that quality demand is strongly correlated with household income.

equilibrium, the number of firms rises until profits are zero and the free entry-condition is satisfied.¹³

The elasticity of the markup with respect to productivity is:

$$\frac{d\nu_t}{\nu_t} = \frac{A\Psi}{1 + A\Psi} \frac{dA_t}{A_t}.$$

This elasticity approaches zero as the fixed cost Ψ approaches zero. For low values of Ψ the model implies that markups are mildly procyclical.

Now consider an increase in θ_t . This shock leads to a fall in the supply of labor, in real income, and in the number of firms that produce differentiated goods. But the markup remains constant.

In sum, the model implies that markups are mildly procyclical. They do not respond to labor supply shocks and are procyclical with respect to changes in productivity. The model is consistent with dispersion in markups across regions. Regions with higher incomes driven by higher productivity choose higher quality goods and pay higher markups.

A natural way to introduce nominal rigidities in this model is to assume that wages are sticky and that the firm has to pay a cost to change the items in its assortment. During recessions it might be optimal for the firm to keep most of its assortment. This sticky assortment is likely to amplify the effect of recessions by limiting the extent to which households can reduce the quality of what they buy.¹⁴ In the time series, we would observe stability in assortment, price and gross margins. In the cross section, we would observe differences in assortment and in markups resulting from the fact that cross sectional differences in income are large and permanent.

6 Conclusion

In this paper we provide direct evidence on the behavior of markups in the retail sector over space and time. We find that gross margins are relatively stable over time and mildly procyclical. At the same time, there is a large regional dispersion of gross margins. Regions

¹³A secular rise in A_t produces a secular increase in the markup that is consistent with the findings of De Loecker and Eeckout (2017).

¹⁴See, Jaimovich, Rebelo and Wong (2015) for an analysis of quality choices during recessions.

with higher incomes consume a different assortment of goods from poorer region and pay higher markups.

We study an endogenous assortment model that is consistent with these basic facts. This model embodies a central insight from the trade model proposed by Fajgelbaum, Grossman and Helpman (2011): higher quality consumption bundles are made of less substitutable components.

7 References

Anderson, Eric, Benjamin A. Malin, Emi Nakamura, Duncan Simester, and Jón Steinsson “Informational Rigidities and the Stickiness of Temporary Sales,” *Journal of Monetary Economics*, 90, 64-83, October 2017.

Anderson, Eric. Nir Jaimovich and Duncan Simester “Price Stickiness: Empirical Evidence of the Menu Cost Channel,” *Review of Economics and Statistics*, vol. 97, issue 4, 813-826, October 2015.

Bertoletti, Paolo and Federico Etro, “Monopolistic Competition when Income Matters,” *The Economic Journal*, 127, 1468-0297, 2017.

Bils, Mark “The Cyclical Behavior of Marginal Cost and Price,” *American Economic Review*, 77(5): 838–55, 1987.

Bils, Mark, and James A. Kahn “What Inventory Behavior Tells Us About Business Cycles.” *American Economic Review*, 90(3): 458–81, 2000.

Bils, Mark, and Peter J. Klenow “Using Consumer Theory to Test Competing Business Cycle Models.” *Journal of Political Economy*, 106(2): 233–61, 1998.

Bils, Mark, and Peter J. Klenow “Quantifying Quality Growth, ” *The American Economic Review*, 91(4): 1006-1030, September 2011.

Bils, Mark, and Peter J. Klenow “Some Evidence on the Importance of Sticky Prices.” *Journal of Political Economy*, 112(5): 947–985, 2004.

Bils, Mark, Peter J. Klenow, and Benjamin A. Malin “Testing for Keynesian Labor Demand.” In *NBER Macroeconomics Annual 2012*, ed. Daron Acemoglu, Jonathan Parker

and Michael Woodford, 311–49. University of Chicago Press, 2012.

Bils, Mark, Peter J. Klenow, and Benjamin A. Malin “Resurrecting the Role of the Product Market Wedge in Recessions,” forthcoming, *American Economic Review*, 2018.

Bloom, Nicholas, Fatih Guvenen and Sergio Salgado “Skewed Business Cycles,” manuscript, University of Minnesota, 2016.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans “Nominal Rigidities and the Dynamic Effects of A Shock to Monetary Policy,” *Journal of Political Economy*, 113(1): 1–45, 2005.

Christiano, Lawrence J., Martin Eichenbaum, and Mathias Trabandt, “Unemployment and Business Cycles,” *Econometrica*, 84 (4), 1523–1569, 2016.

Cochrane, John, and Monika Piazzesi, “The Fed and Interest Rates: A High-Frequency Identification,” *American Economic Review*, 92(2), 2005.

Collard-Wexler, Allan and Jan De Loecker “Reallocation and Technology: Evidence from the US Steel Industry,” *American Economic Review*, Vol. 105, No. 1, pp. 131-171, January 2015.

De Loecker, Jan and Jan Eeckhout “The rise of market power and the macroeconomic implications,” *National Bureau of Economic Research*, Working Paper No. 23687, 2017.

Della Vigna, Stefano and Matthew Gentzkow “Uniform Pricing in U.S. Retail Chains,” manuscript, U.C. Berkeley, 2017.

Eichenbaum, Martin, Nir Jaimovich and Sergio Rebelo, “Reference Prices, Costs, and Nominal Rigidities,” *American Economic Review*, Vol. 101, No. 1, pp. 234-262, February 2011.

Erceg, Christopher J., Dale W. Henderson, and Andrew T. Levin “Optimal Monetary Policy with Staggered Wage and Price Contracts,” *Journal of Monetary Economics*, vol. 46, no. 2, 281-314, 2000.

Fajgelbaum, Pablo, Gene M. Grossman and Elhanan Helpman “Income Distribution, Product Quality, and International Trade,” *Journal of Political Economy*, Vol. 119, No. 4, pp. 721-765, August 2011.

Galeotti, M., and E Schiantarelli, “The Cyclicalities of Markups in a Model with Adjustment Costs: Econometric Evidence for U.S. Industry,” *Oxford Bulletin of Economics and Statistics* 60:121-142, 1998.

Coibion, Olivier, Yuriy Gorodnichenko, and Gee Hee Hong, “The Cyclicalities of Sales, Regular and Effective Prices: Business Cycle and Policy Implications,” *The American Economic Review* 105 (3): 993-1029.

Golosov, Mikhail and Robert E. Lucas Jr., “Menu Costs and Phillips Curves,” *Journal of Political Economy*, University of Chicago Press, vol. 115, pages 171-199.

Gorodnichenko, Yuriy, and Michael Weber, “Are Sticky Prices Costly? Evidence from the Stock Market,” *American Economic Review* 106(1): 165-99.

Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson “Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements,” *International Journal of Central Banking*, 1(1): 55–93, 2005.

Hall, Robert E. “The Relation between Price and Marginal Cost in U.S. Industry.” *Journal of Political Economy* 96 (October): 921–47, 1988.

Hall, Robert E. “The Cyclical Response of Advertising Refutes Counter-Cyclical Profit Margins in Favor of Product Market Frictions.” National Bureau of Economic Research Working Paper 18370, Cambridge, MA., 2012.

Jaimovich, Nir and Max Floetotto “Firm Dynamics, Markup Variations, and the Business Cycle,” *Journal of Monetary Economics*, Volume 55, Issue 7, Pages 1238-1252, October 2008.

Jaimovich, Nir Sergio Rebelo and Arlene Wong “Trading Down and the Business Cycle,” National Bureau of Economic Research Working Paper 21539, Cambridge, MA., 2015.

Karabarbounis, Loukas “The Labor Wedge: MRS vs. MPN,” *Review of Economic Dynamics*, 17(2), 206223.

Klenow, Pete and Benjamin Malin “Microeconomic Evidence on Price-Setting,” *Handbook of Monetary Economics*, 3A, B. Friedman and M. Woodford ed.: Elsevier, 2011, 231-284.

Kuttner, Kenneth “Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market,” *Journal of Monetary Economics*, 47: 523–544, 2001.

Midrigan, Virgiliu “Menu Costs, Multiproduct Firms, and Aggregate Fluctuations,” *Econometrica*, 79(4), pages 1139-1180, July 2011.

Nakamura, Emi and Jón Steinsson “Monetary Non-Neutrality in a Multisector Menu Cost Model,” *Quarterly Journal of Economics*, pages 961-1013, August 2010.

Nakamura, Emi and Jón Steinsson “High Frequency Identification of Monetary Non-Neutrality: The Information Effect,” *Quarterly Journal of Economics*, forthcoming, 2018.

Neiman, Brent and Joseph Vavra “The Rise in Household Spending Concentration,” Working Paper, SSRN, March 2018.

Nekarda, Christopher and Valerie A. Ramey, “The Cyclicalty of the Price-Cost Markup,” manuscript, University of California at San Diego, November 2013.

Pasten, Ernesto, Raphael Schoenle, Michael Weber, “Price Rigidities and the Granular Origins of Aggregate Fluctuations,” NBER Working Paper No. 23750, August 2017.

Ramey, Valerie A. and Daniel J. Vine “Oil, Automobiles, and the U.S. Economy: How Much Have Things Really Changed?,” *NBER Macroeconomics Annual*, Vol. 25, No. 1, pp. 333-368, 2010.

Ravn, Morten Stephanie Schmitt-Grohé, and Martín Uribe “Deep Habits,” *The Review of Economic Studies*, Volume 73, Issue 1, Pages 195–218, January 2006.

Rotemberg, Julio J., and Michael Woodford “The Cyclical Behavior of Prices and Costs.” In *Handbook of Macroeconomics.* , ed. John B. Taylor and Michael Woodford, 1051–135. Elsevier, 1999.

Saiz, Albert “The Geographical Determinants of Housing Supply,” *The Quarterly Journal of Economics*, Vol. 125, No. 3, pp. 1253-1296, August 2010.

Stroebel, Johannes and Joseph Vavra “House Prices, Local Demand and Retail Prices,” forthcoming, *Journal of Political Economy*, 2018.

Woodford, Michael *Interest and Prices: Foundations of A Theory of Monetary Policy*. Princeton, N.J.:Princeton University Press, 2003.

A Appendix

A.1 Monetary policy and oil shocks

In section 3.2, we study the conditional response of firms' gross and net operating margins to high-frequency monetary policy shocks and oil-price shocks. This appendix discusses how these shocks are identified.

Monetary policy shocks are identified using high-frequency data on the Federal Funds futures contracts. This approach has been used by Kuttner (2001), Cochrane and Piazzesi (2002), Nakamura and Steinsson (2018), Gorodnichenko and Weber (2015), and others. The future rate reflects the market expectations of the average effective Federal Funds rate during that month. It therefore provides a market-based measure of the anticipated path of the Federal Funds rate.

A current period monetary policy shock is defined as:

$$\epsilon_t = \frac{D}{D-t} (ff_{t+\Delta+}^0 - ff_{t-\Delta-}^0) \quad (11)$$

where t is the time when the FOMC issues an announcement, $ff_{t+\Delta+}^0$ is the Federal Funds futures rate shortly after t , $ff_{t-\Delta-}^0$ is the Federal Funds futures rate just before t , and D is the number of days in the month. The $D/(D-t)$ term adjusts for the fact that the Federal Funds futures settle on the average effective overnight Federal Funds rate.

We consider a 60-minute time window around the announcement that starts $\Delta^- = 15$ minutes before the announcement. Examining a narrow window around the announcement ensures that the only relevant shock during that time period (if any) is the monetary policy shock. Following Cochrane and Piazzesi (2002) and others, we aggregate up the identified shocks to obtain a quarterly measure of the monetary policy shock.

Oil-price shocks are identified using the approach proposed by Ramey and Vine (2010), updated to the recent period. We estimate a VAR system with monthly data

$$Y_t = A(L)Y_{t-1} + U_t.$$

The vector Y_t includes the following variables (in order): nominal price of oil, the CPI, nominal wages of private production workers, industrial production, civilian hours, and the

federal funds rates. The function $A(L)$ is a matrix of polynomials in the lag operator L , and U is a vector of disturbances. All variables, except the federal funds rate, are in logs. We include a linear time trend and 6 lags of the variables. The shock to oil prices is identified using a standard Cholesky decomposition. The shocks are aggregated to a quarterly frequency to match the frequency of our firm level data.

A.2 Correcting gross margins for changes in inventories

One potential source of measurement error in our aggregate retail and firm level data stems from the possibility that the cost of goods sold might reflect goods purchased in previous periods and stored as inventory. As a result, the cost of goods sold does not measure the true marginal replacement cost.

We deal with this issue in Section 3.4 by using actual replacement cost for a retailer. Here, we use instead a perpetual inventory approach to correct the cost of goods sold for changes in inventories.

Denote by \bar{C}_t the observed cost of goods sold and by C_t the true cost of goods sold. The observed cost of goods sold is

$$\bar{C}_t = \alpha_t \bar{C}_{t-1} + (1 - \alpha_t) C_t,$$

where

$$\alpha_t = \frac{\text{Starting period inventories}_t}{\text{Sales}_t}.$$

We assume that if $\alpha_t \geq 1$, then

$$\bar{C}_t = C_t / (1 + \pi_t),$$

where π_t is the rate of change in the producer price index for final goods from the Bureau of Labor Statistics. This equation implies that, if the inventories at the start of the period exceed sales in that period, then the goods sold in that period come from inventories.¹⁵ The observed value of cost of good sold is then assumed to be given by the true cost of goods sold, deflated by the producer price index.

¹⁵This occurrence is rare, particularly at the annual frequency. The average retailer ratio of inventories to sales is about 12%.

The true cost of goods sold is given by

$$C_t = \frac{\bar{C}_t - \alpha_t \bar{C}_{t-1}}{1 - \alpha_t}, \quad \text{if } \alpha_t < 1$$

and

$$\bar{C}_t = C_t / (1 + \pi_t), \quad \text{if } \alpha_t \geq 1.$$

We assume as starting value $\bar{C}_0 = C_0$ and implement our approach separately for each firm.

The gross margin adjusted for changes in inventories is given by

$$\frac{\text{Sales}_t - C_t}{\text{Sales}_t}.$$

We use this adjusted measure to re-estimate the elasticity of gross margins with respect to real GDP. We regress the year-on-year logarithmic difference of each variable on the year-on-year logarithmic difference of real GDP.

Table 10 shows our results from Section 3, which do not adjust for inventories, as well as the elasticities estimated using gross margins adjusted for changes in inventories. We see that while point estimates are different, the elasticity of gross margins with respect to GDP growth remains statistically insignificant when we use the adjusted measures of gross margins.

A.3 Proof of proposition 1

Equilibrium in the homogeneous good market requires:

$$w_t = (1 + g)^t.$$

The equilibrium price index for consumption of differentiated goods is:

$$P_t = v_t^{-\gamma} n_t^{1-v_t} \frac{\nu_t}{A_t}.$$

Households choose the same consumption level for all available varieties: $x_{ivt} = x_{vt}$. The consumption bundle is given by:

$$C_t = \nu_t^\gamma n_t^{v_t} x_{vt}.$$

Using the household budget constraint, we can rewrite C_t as:

$$C_t = \nu_t^{\gamma-1} n_t^{v-1} \alpha Y_t A_t.$$

Since $\gamma > 1$, the household's utility is monotonically increasing in v_t . The value of x_{vt} is given by:

$$x_{vt} = \frac{\alpha A_t Y_t}{\nu_t n_t}. \quad (12)$$

Since utility is increasing in ν_t , the constraint $x_{vt} \geq 1$ is binding. Setting $x_{vt} = 1$ in equation (12), we obtain the optimal value of ν_t :

$$\nu_t = \frac{\alpha A_t Y_t}{n_t}. \quad (13)$$

The monopolist profits are equal to:

$$\pi_t = \frac{1}{A_t} (\nu_t - 1) - \Psi.$$

The free entry condition, $\pi_t = 0$, implies that the markup is given by:

$$\nu_t = 1 + \Psi A_t. \quad (14)$$

Using equation (13) to replace ν_t , we obtain:

$$n_t = \frac{\alpha A_t Y_t}{1 + \Psi A_t}. \quad (15)$$

Equilibrium prices are given by:

$$p_{ivt} = \frac{v_t}{A} = \Psi + 1/A_t.$$

Household income is given by:

$$Y_t = (1 + g_t)^t N_t. \quad (16)$$

The equilibrium value of N_t is given by:

$$N_t = N = \frac{1}{1 + \theta_t}. \quad (17)$$

Combining equations (15), (16), and (17), we obtain the following expression for the equilibrium number of monopolistic firms:

$$n_t = \frac{\alpha A_t N (1 + g_t)^t}{1 + A_t \Psi}. \quad (18)$$

To solve for real income, we replace Y_t in equation (8):

$$\tilde{Y}_t = \frac{1}{P_t^\alpha} \frac{1}{1 + \theta_t} (1 + g)^t$$

Replacing P_t :

$$\tilde{Y}_t = \frac{A_t^\alpha \left(\frac{\alpha}{\Psi+1/A_t} \frac{1}{1+\theta_t} \right)^{\alpha(v-1)}}{v_t^{(1-\gamma)\alpha}} \frac{1}{1+\theta_t} [(1+g)^t]^{1-\alpha(1-v_t)}.$$

Using the fact that:

$$1 - \nu_t = -\Psi A_t,$$

we obtain,

$$\tilde{Y}_t = \frac{A_t^\alpha \left(\frac{\alpha}{\Psi+1/A_t} \frac{1}{1+\theta_t} \right)^{\alpha\Psi A_t}}{(1+\Psi A_t)^{(1-\gamma)\alpha}} \frac{1}{1+\theta_t} [(1+g)^t]^{1+\alpha\Psi A_t}.$$

To see that \tilde{Y}_t is an increasing function of A_t , it is convenient to take logarithms:

$$\begin{aligned} \log(\tilde{Y}_t) &= \alpha \log(A_t) + \alpha\Psi A_t \log\left(\frac{\alpha}{\Psi+1/A_t} \frac{1}{1+\theta_t}\right) + (\gamma-1)\alpha \log(1+\Psi A_t) - \log(1+\theta_t) \\ &\quad + [1+\alpha\Psi A_t] t \log(1+g). \end{aligned}$$

Tables and Graphs

Table 1: Cyclicalities of Aggregate Retail Trade Variables

	Elasticity wrt GDP			
	Quarterly		Annual	
Gross margins	0.162	(0.256)	0.376	(0.616)
Operating profit margins	2.286**	(0.895)	5.233	(3.632)
Sales	8.089***	(0.45)	9.279***	(1.976)
Cost of goods sold	8.104***	(0.43)	9.140***	(2.154)

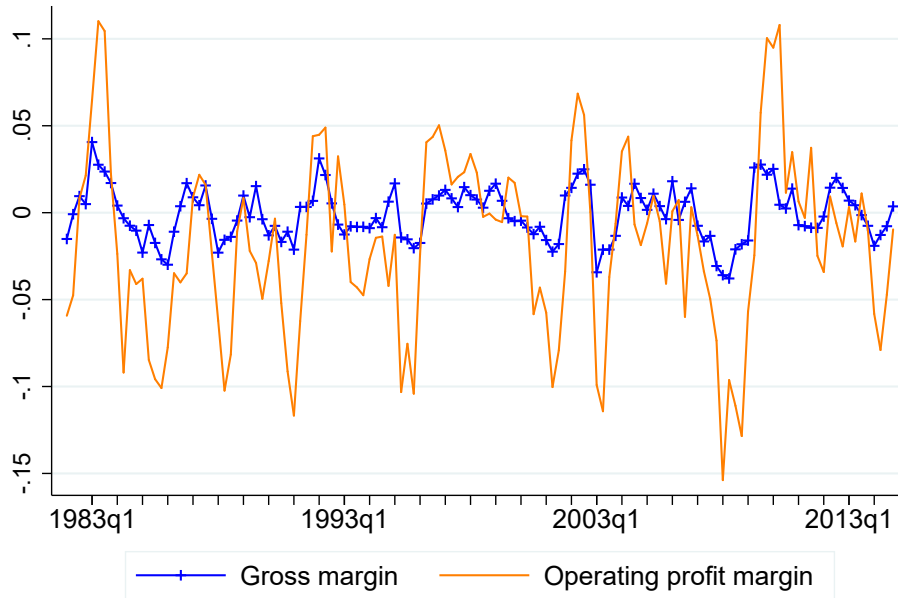
Notes: Variables are log-difference from prior year. Data is from Compustat and the BLS. Each row is estimated from a separate regression of the variables on GDP. We estimate the elasticities at quarterly and annual frequencies. See text for more details. Standard errors are in parentheses. *, **, and *** give the significance at the 10, 5, and 1 percent levels.

Table 2: Volatility of Aggregate Retail Trade Variables

Standard Deviation	Quarterly	Annual
Gross margins	0.017	0.011
Operating profit margins	0.057	0.051
Sales	0.046	0.062
Cost of goods sold	0.045	0.060

Notes: Variables are log-difference from prior year. Data is from Compustat and the BLS. The standard deviations are computed at quarterly and annual frequencies. See text for more details.

Figure 1: Time-series of Aggregate Retail Trade Variables



Notes: Variables are log-difference from prior year. Data is from Compustat and the BLS. The data is plotted at a quarterly frequency.

Table 3: Cyclicity of Firm-Level Variables

	Elasticity wrt GDP			
	Quarterly		Annual	
Gross margins	0.31	(0.37)	0.15	(0.55)
Operating profit margins	3.03***	(0.96)	3.60***	(1.11)
Sales	3.18***	(0.32)	3.64***	(0.67)
Cost of goods sold	3.09***	(0.32)	3.58***	(0.70)

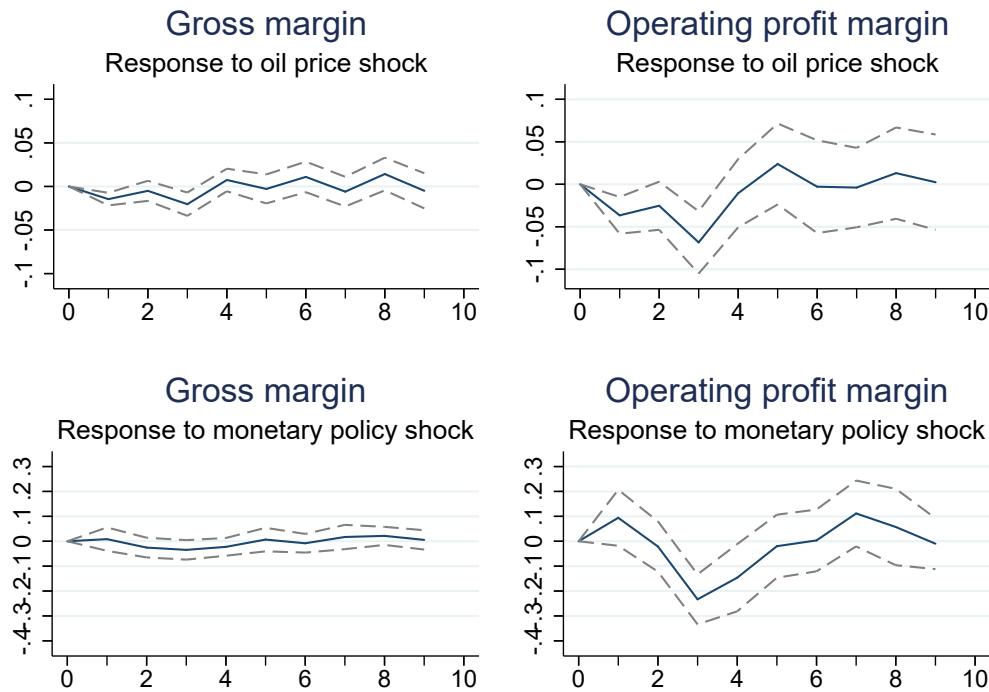
Notes: Variables are log-difference from prior year. Data is from Compustat and the BLS. Each row is estimated from a separate regression of the variables on GDP. We estimate the elasticities at quarterly and annual frequencies. See text for more details. Standard errors are in parentheses. *, **, and *** give the significance at the 10, 5, and 1 percent levels.

Table 4: Volatility of Firm-Level Variables

Standard Deviation	Quarterly	Annual
Gross margins	0.061	0.480
Operating profit margins	0.254	0.699
Sales	0.080	0.364
Cost of goods sold	0.084	0.407

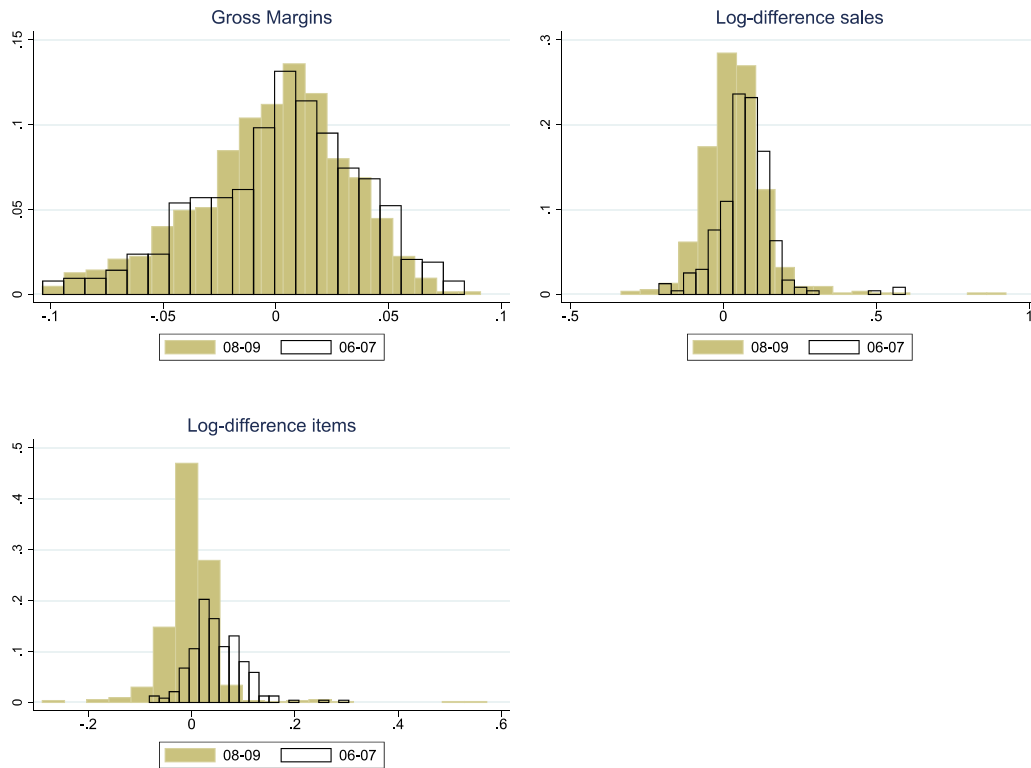
Notes: Variables are log-difference from prior year. Data is from Compustat and the BLS. The standard deviations are computed at quarterly and annual frequencies. See text for more details.

Figure 2: Impulse Response Functions to Monetary Policy and Oil Price Shocks



Notes: The figure depicts the impulse response functions of the (log-differenced) gross margins and net operating profit margins to a 1ppt monetary policy shock (bottom panel) and an oil price shock (top panel). See text for more information. The data is plotted at a quarterly frequency. Dashed lines are the 90th percentile to a 1 ppt shock.

Figure 3: Histograms of Gross Margins, Sales, and Number of Items



Notes: The figure depicts the distributions of average gross margins (levels), sales (log-difference from same quarter in the prior year) and number of items (log difference from same quarter in the prior year) for the period 2006-07 and the period 2008-09. See text for more details. For confidentiality purposes, we normalize the distribution of gross margin by the mean margin in 2006-07. Specifically, we subtract the mean 2006-07 margin from the 2006-07 distribution, so at the mean it is zero. We also subtract the mean 2006-07 margin from the 2007-08 distribution.

Table 5: Cross-sectional Distribution of Margins, Sales and Number of Items

	Mean	p10	p50	p90
Margins (levels)				
Difference	-0.005	-0.003	-0.003	-0.007
Log difference in sales				
2006-07	0.072	-0.026	0.072	0.154
2008-09	0.038	-0.074	0.034	0.145
Difference	-0.034	-0.048	-0.037	-0.009
Log difference in number of items				
2006-07	0.050	-0.007	0.044	0.111
2008-09	0.000	-0.053	-0.001	0.043
Difference	-0.050	-0.046	-0.045	-0.068

Notes: Data is from a large retailer. The table gives key moments from the cross-sectional distribution (across regions) of gross margins, average sales growth and average growth in number of items. We report the average levels of each variable in 2006-07 and 2008-09, and the differences between 2006-07 and 2008-09 for sales growth and growth in number of items. Due to confidentiality reasons, we do not report the *levels* of the margins, and only report how the level of margins changed between 2006-07 and 2008-09.

Table 6: Cyclicity of Store-Item Variables

	Elasticity with respect to local UR		Elasticity with respect to local house prices	
Gross margin	0.021	(0.026)	0.075**	(0.03)
Price	-1.465*	(1.206)	0.10	(0.07)
Replacement cost	-0.358	(0.638)	0.021	(0.09)
Sales	-0.902**	(0.36)	0.249***	(0.09)
Number of items	-1.057***	(0.02)	0.208***	(0.07)

Notes: Variables are log-difference from prior year. Data is from a large retailer. Each entry is a separate regression of the log-differenced variable on the local area change in unemployment rate and house prices. Standard errors are clustered by county. See text for more details.

Table 7: Volatility of Store-Item Variables

	Stdev
Markup	0.026
Price	0.009
Replacement cost	0.005
Sales	0.220
Number of items	0.118

Notes: Variables are log-difference from prior year. Data is from a large retailer. The standard deviations are computed at a quarterly frequency. See text for more details.

Table 8: Variance Decomposition of the Cross-sectional Margins

Spatial variation due to:	Item sold everywhere	Item not sold everywhere
(i) Differences in gross margins for the same item	10%	3%
(ii) Differences in assortment composition	85%	81%
(iii) Interaction term	1%	1%
(iv) Covariance term	4%	15%

Notes: Data is from a large retailer. The table gives the decomposition of the cross-sectional variance (across regions) into the four components: differences in gross margins for the same item, differences in assortment of composition, the interaction terms, and the covariance terms. See text for more details.

Table 9: Cross-sectional Variation in Margins and Regional Characteristics

	Estimate	Std error
Log household income	0.170***	(0.064)
Log median house value	0.161***	(0.055)
Share of income to top 1%	0.707***	(0.179)
Herfindahl index	-0.009	(0.048)
Rural county	0.025	(0.008)

Notes: Table gives the elasticity of the gross margin with respect to each of the variables. Each regression is estimated separately. Standard errors are in parentheses. *, **, and *** give the significance at the 10, 5, and 1 percent levels.

Table 10: Appendix A.2: Cyclicalities of Gross Margins, Adjusting for Inventories

	Gross Margin Elasticity wrt GDP			
	Quarterly		Annual	
Regressions: baseline				
Industry-level regression	0.162	(0.256)	0.376	(0.616)
Firm-level regression	0.310	(0.373)	0.152	(0.548)
Regressions: with inventory adjustment				
Industry-level regression	-0.231	(1.45)	-0.522	(0.924)
Firm-level regression	-0.550	(2.647)	-0.351	(0.678)

Notes: Table gives the elasticity of the gross margin with respect to each of the variables. Each regression is estimated separately. Standard errors are in parentheses. *, **, and *** give the significance at the 10, 5, and 1 percent levels. The baseline regressions are from Section 3 and correspond to the estimates from Table 1. The regression estimates with inventory adjustment are based on the perpetual inventory approach. See Appendix A2 for details.