The Changing Anatomy of Firms' Cost Structure in the US*

Hassan Afrouzi[†] Columbia
Andres Drenik[‡] Columbia
Ryan Kim[§] Johns Hopkins

February 15, 2020
preliminary and incomplete

Abstract

Recent literature documents that the share of firms' costs devoted to production (COGS-to-OPEX ratio) has declined over the past several decades, which has also been related to the rise in markups in the US. We document that the decline in the share of firms' costs devoted to production is mainly driven by entrants in industries closely related to final consumers. Moreover, we document that the rising component of firms' costs is highly and positively correlated with firms' sales, as well as recurring consumers, which indicates that these costs are variable. We develop a quantitative model to study the implications of these findings for the measurement of markups. Within our framework, while price over cost of production markups increase with size and productivity, the relationship is much smaller for price over cost of of goods sold.

JEL classification: E2, D2, D4

Key Words: Cost of Goods Sold, Operational Costs, Markups, Inertial Demand

*We thank Luigi Caloi for providing superb research assistance. We also thank John Asker and Virgiliu Midrigan for valuable comments and suggestions. Researchers own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[†] hassan.afrouzi@columbia.edu.
[‡] ad3376@columbia.edu
[§] rkim59@jhu.edu

1 Introduction

Recent evidence suggests that price over cost of production markups have increased over the last several decades in the U.S. (De Loecker et al., 2018; Edmond et al., 2018).[1] However, neither the underlying causes nor the welfare implications of this trend are well understood. In particular, it can either be driven by an increase in market power, or changes in the composition firms' costs. For instance, if firms have to pay a larger costs to attain the same market share, then higher price over cost of production markups are not necessarily indicative of higher market power. Traina (2019) provides evidence for this channel by documenting that the share of firms' operational costs that are allocated to producing goods and services (hereinafter COGS-to-OPEX ratio) has declined steadily starting in the 1960s. In this paper, we study three closely connected questions: (1) Which firms are the main drivers of the fall in the average COGS-to-OPEX ratio? (2) Why is it falling? and (3) What are the implications of this decline for the rise of price over cost of production markups?

The starting point of our analysis is to characterize the dynamics of firms' cost structure using micro-data from Compustat, which contains data from financial statemets of public firms in the US since the 1960s. The average firm's COGS as a share of their OPEX has declined steadily from 83% in 1960s to 69% in the past decade. The goal is to understand whether the decline is a compositional phenomenon or a within-firm trend over time. To study this, we decompose the decline in the COGS-to-OPEX ratio to a component that comes from incumbent firms (i.e., the intensive margin) and two other components that capture the difference between entrants' and exiters' COGS-to-OPEX ratio relative to incumbents (i.e., the extensive margin). We find that the extensive margin is the sole driver of the decline since the 1980s and explains over 57% of the decline over the whole sample (starting in the 1960s).

On the extensive margin, both entrants and exiters have a lower COGS-to-OPEX ratio than incumbents, and the selection among exiters is mainly driven by those that exit shortly after entry. These findings suggest that entrants are the main drivers of this decline over the sample. A natural follow-up question is whether this selection among entrants is due to different trends of entry in different industries. We find that this is purely a within-industry phenomenon

The second contribution of this paper is to provide evidence about potential mechanisms driving the decline in the COGS share of OPEX. We do this in two steps. First, we be-

[1] De Loecker et al. (2018) document an increase in sales-weighted markups and relate it to a rise in market power. Edmond et al. (2018) document a mitigated increase in cost-weighted markups and quantify the cost of markups (see also Gutierrez et al., 2019; Weiss, 2019; Zhang, 2019). Moreover, in a broader context, complementary evidence on the rise of superstar firms (David et al., 2017) or the decline of the labor share of income (Kehrig and Vincent, 2018; Gouin-Bonenfant, 2018) are also consistent with this trend.

gin by analyzing the characteristics of firms that are driving the decline in COGS-to-OPEX ratio. We show that there are differential effects across entrants that are "closer" to final consumers, which is measure by an industry's average share of output absorbed by final consumers. Furthermore, the trend is more pronounced for firms whose expenditures on advertising relative to sales are above the median. Accordingly, we conclude that the decline in the COGS share of OPEX seems to be driven by young firms' that have a stronger relationships with final consumers.

Motivated by these results, we study changes in consumers' behavior in more detail. The previous literature has identified trends on the consumer side that could be potentially relevant to this decline. Bornstein (2018), for instance, documents that persistence of households' consumption choices has increased over time in the U.S. and provides a framework that shows this phenomenon is consistent with a decline in the share of young firms and higher profit share of income. Similarly, Neiman and Vavra (2019) document that the average household has increasingly concentrated its spending on a few preferred products. We provide additional evidence on changes in consumers' willingness to change their consumption patterns. Using ACNielsen Homescan Panel data, we show that, in addition to these documented trends, the number of stores that households visit to make their purchases has declined over time as well. We interpret these trends as evidence for a lower propensity of switching to new products on the side of consumers.

More importantly, by merging the Compustat with the ACNielsen data, we analyze how the probability of a household to purchase the same good over time is related to the firm's cost structure. We find a strong positive relationship between a firm's spending on SG&A expenses and this probability.

A lower propensity of switching among households would explain why young firms would need to spend more in a category other than the cost of goods sold: they need to spend more than their predecessors to gain the same market share, as their consumers nowadays are less likely to switch. Furthermore, it is important to note that the higher cost of attracting new consumers, which manifests itself as a decline in the COGS share of OPEX in the data, is not necessarily a rise in market power or even a rise in the share of fixed costs. Firms' costs that are aimed at attracting new customers could potentially have a variable cost nature, which is proportional to the level of production but does not fit the classical definition of a cost *for* production. In fact, we find that the increasing component of firms' costs in the data is strongly correlated and proportional to firms' sales, which is more consistent with the definition of a variable cost than a fixed cost.

Literature review. [TBC] The project is related to a recent literature measuring the trends in market power in the U.S. De Loecker et al. (2018) document an increase in sales-weighted markups and relate it to a rise in market power. Edmond et al. (2018) document a mitigated increase in cost-weighted markups and quantify the cost of markups (see also Gutierrez et al., 2019; Weiss, 2019; Zhang, 2019). Traina (2019) documents an aggregate decline in the COGS-to-OPEX ratio and shows the impact this decline has for the measurement of markups. Another strand of the literature has identified trends on the consumer side that could be potentially relevant to this decline. Bornstein (2018) documents that persistence of households' consumption choices has increased over time in the U.S. Similarly, Neiman and Vavra (2019) document that the average household has increasingly concentrated its spending on a few preferred products. We contribute to these literatures by providing a thorough characterization of changes in firms' cost structures, a description of its potential determinants, and with the help of a model an exploration of how these changes affect the measurement of markups and optimal policies.

2 Empirical Facts

In this section, we present our main fact regarding the secular decline of the COGS-to-OPEX ratio, and the predominat role played by entrant firms that operate in industries "close " to final consumers.

2.1 Data Description

For the analysis of firms' cost structure, we draw firm-level financial micro-data from Compustat, which includes panel data on publicly traded firms since 1960. The most important feature of Compustat for our analysis is the availability of detailed information on firms' balance sheets. More specifically, it contains information on a firm's cost structure. With the caveat that it only covers public firms, it still constitutes the main source of data for firm-level analysis in the US and has been used by the recent literature on firm markups (see, for example, De Loecker et al., 2018; Edmond et al., 2018; Traina, 2019).

Throughout the analysis, we focus on two measures of a firm's costs from Compustat. From an accounting perspective, a firm's costs associated with the running of the firm are captured in the Operating Expense (OPEX) account. In turn, OPEX is divided into two types of costs: Cost of Goods Sold (COGS) and Selling, General and Administrative Expenses (SGA). According to Compustat, COGS captures *"all expenses that are directly related to the cost of merchandise purchased or the cost of goods manufactured that are withdrawn from finished goods inventory and sold to customers"*. That is, it records costs attributable to the production of the goods sold by a firm. Typical categories included are the cost of labor used in production and the costs of intermediate inputs. On the other hand, SGA captures *"all com-*

mercial expenses of operation (such as expenses not directly related to product production) incurred in the regular course of business...". In other words, it includes the costs incurred to sell and deliver products and services and the costs to manage the company. Typical categories included within SGA are advertising, delivery, marketing, research and development, among others. The main variable of interest is the COGS-to-OPEX ratio, that is, the share of operating expenses devoted to the production of a firm's goods or services. In Appendix A, we provide a detailed description of the variables used and the cleaning procedure.

When providing evidence of potential mechanism behind the decline in the COGS-to-OPEX ratio, we use the ACNielsen Homescan Panel, which was made available by the Kilts Marketing Data Center at the University of Chicago Booth School of Business. The data contain approximately 4.5 million barcode-level product sales recorded daily from an average of 55,000 households per year in the United States. A barcode is a unique universal product code (UPC) assigned to each product and is used to scan and store product information. All households sampled by Nielsen are provided with in-home scanners to record their purchases of products with barcodes. Nielsen assigns a sample weight—or a projection factor—to each household based on 10 demographic variables to make the sample nationally representative.[2] The data record information on which retailer each household visited to purchase products at a given point of time. According to Nielsen, the Homescan Panel covers approximately 30 percent of all household expenditures on goods in the consumer price index (CPI) basket. The data we use cover the period of 2004-2016.

In addition, we merge the Nielsen database with the Compustat data by using the GS1 US Data Hub. GS1 is the company that issues barcodes to producers.[3] Their data record the company name for each barcode-level product, providing a way to link customer behavior and barcode-level product information with its producer information. A "firm" in the database is defined based on the entity that purchased the barcodes from the GS1. Thus, a firm in the data could be a manufacturer, such as a Coca-Cola manufacturer that needs a barcode to sell its cherry-flavored 500-ml diet coke, or it could be a small retailer that wants to sell its private-label products. We further combine the producer information with the Compustat for the period of 2004-2017.[4] In using the Nielsen data, we sometimes divide

[2]The 10 demographic variables are: household size, household income, head of household age, race, Hispanic origin, male head education, female head education, head of household occupation, the presence of children, and Nielsen county size.

[3]GS1 provides a business with up to 10 barcodes for a $250 initial membership fee and a $50 annual fee. There are significant discounts on the cost per barcode for firms purchasing larger quantities of barcodes (see http://www.gs1us.org/get-started/im-new-to-gs1-us).

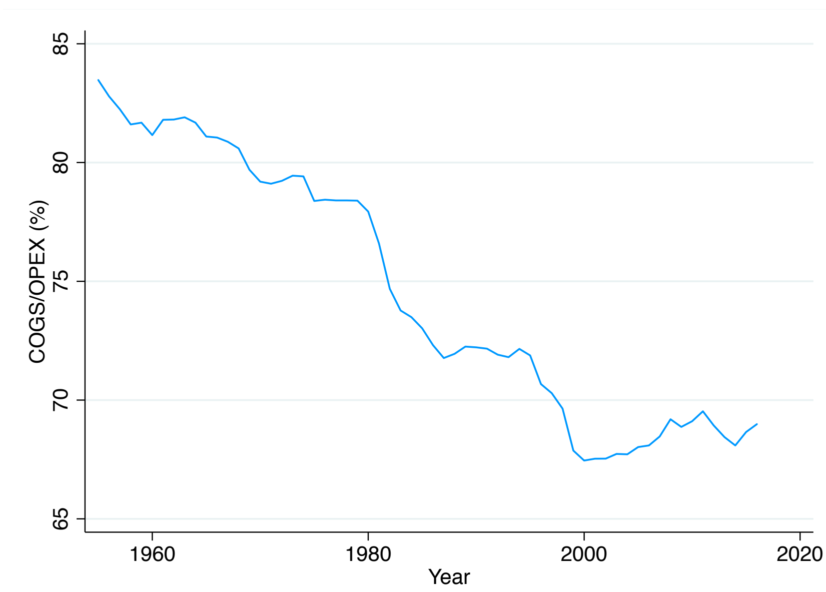
[4]To match two different databases, we use the "reclink" STATA software command based on company name after standardizing it with the "std_compname" command. Once Stata reports the results, we only pick those companies that have higher than .99 threshold. Then we manually check the company name for every company in this threshold and drop those that look suspicious

our sample into categories of products (product group) and locations where households live and purchase products (Scantrack market). The product group is a category of products, such as pet food and school supplies. The Scantrack market is the geographic region defined in Nielsen and is similar to that of the metropolitan statistical area. In aggregating all the measures, we use the household sample weight and their total spending on products available in Nielsen data.

2.2 The Secular Decline in the COGS-to-OPEX Ratio

Aggregate Decline. We document the change in firms' cost structures by first plotting the aggregate COGS-to-OPEX ratio for the entire period (see Figure 1). Since at this stage the analysis is descriptive, we present results based on unweighted averages. As Traina (2019) documents using Compustat data, the average firm's COGS as a share of their OPEX has declined steadily from 83% in 1960s to 69% in the past decade. Therefore, direct costs of good production are no longer the only major cost of operation for firms, and Selling, General and Administrative expenses (SGA) seem to have grown rapidly over time. Our main objective in this empirical section is to provide a comprehensive characterization of the observed change in firms' cost structure over time, sectors and firms.

Figure 1: COGS Share of OPEX over Time



Notes: The figure shows the average COGS-to-OPEX ratio over time for firms in the Compustat database.

Incumbents, Entrants and Exiters. Our analysis begins by asking: Is the decline in the COGS-to-OPEX ratio driven by the intensive margin or the extensive margin? To answer this question, we decompose the change in the aggregate COGS-to-OPEX ratio to the part coming from within incumbent firms and the parts due to entry and exit:[5]

$$\begin{aligned}\Delta y_t = & \underbrace{\mathbb{E}[\Delta y_{i,t}|\text{incumbents}]}_{\text{growth within incumbents}} \\ & + \underbrace{\left(1 - \frac{s_t}{1+g_t}\right)\left(\mathbb{E}[y_{i,t}|\text{entrants}] - \mathbb{E}[y_{i,t}|\text{incumbents}]\right)}_{\text{selection among entrants}} \\ & + \underbrace{(1 - s_t)\left(\mathbb{E}[y_{i,t-1}|\text{incumbents}] - \mathbb{E}[y_{i,t-1}|\text{exiters}]\right)}_{\text{selection among incumbents}}.\end{aligned} \tag{1}$$

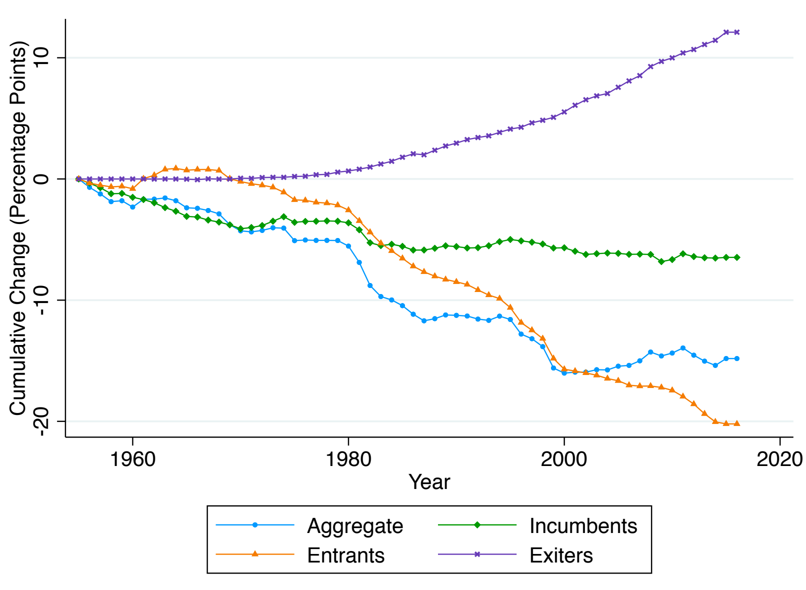
Here, y_t denotes the aggregate COGS-to-OPEX ratio at time t, $y_{i,t}$ is the COGS-to-OPEX ratio of firms i at time t, s_t is the survival rate of firms from time $t-1$ to t, g_t is the growth are of firms' population from $t-1$ to t, and all expectations are taken in the cross-section of firms conditional on the specified group. The decomposition simply states that in addition to growth within incumbents, the aggregate COGS-to-OPEX ratio can also decrease either because entrants enter with a lower COGS-to-OPEX ratio than incumbents, or exiters exit with a higher COGS-to-OPEX ratio than incumbents.

Figure 2 plots the three components of this decomposition for the entire period. We normalize the COGS-to-OPEX ratio to its value in 1955 and plot the cumulative change for each component in percentage points over time. Of the 14 percentage points decline in the aggregate COGS-to-OPEX ratio, around 6 percentage points (43%) is coming from the intensive margin – the change within incumbents – and the remaining 8 percentage points (57%) are due to the extensive margin. Moreover, the analysis also reveals that, on the extensive margin, both entrants and exiters have a lower COGS-to-OPEX ratio than incumbents. In fact, the cumulative decline in the COGS-to-OPEX ratio among entrants is about 20 percentage points, which is larger than the aggregate decline. This sharp decline among entrants is partially compensated by the difference between exiters and incumbents, with a cumulative difference of about 12 percentage points throughout the sample.

More importantly, the Figure shows that the fraction of the decline in the COGS-to-OPEX ratio coming from within incumbents almost entirely happened before the 1980s and has been stagnant ever since. In contrast to within incumbent changes, however, the net decline coming from the two extensive margins *started* in the 1980s and has been going on steadily since then. Therefore, not only the decline coming from the extensive margin explains a

[5]By entry and exit, we refer to a firm's entry into and exit out of the dataset. According to Compustat, entry and exit occurs because of firms' decisions to become public or private, bankruptcy, or mergers and acquisitions.

Figure 2: Decomposition of the COGS Share of OPEX



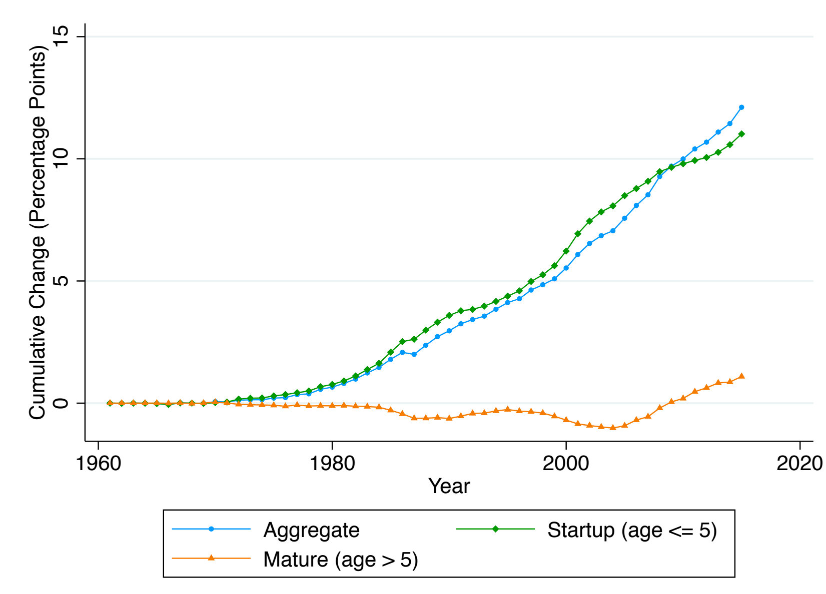
Notes: The figure shows the cumulative change in the COGS-to-OPEX ratio over time for firms in the Compustat database, decomposed to the part coming from within incumbents and the parts due to firms that entry and exit (see equation (1)). We normalized the COGS-to-OPEX ratio to 1955's ratio and the plot illustrates the cumulative change in percentage points over time.

larger share of the aggregate decline, but it also seems to be the main driver of the changes in recent decades.[6]

Thus, the entry and exit margins seem to be of crucial importance in understanding the change in the anatomy of firms' cost structure. It appears that firms enter with lower COGS-to-OPEX ratio in later years, a pattern that is also shared with exiters. One key question is, therefore, who are the exiters? Are they part of the older batch of firms, or are they from the group of recent entrants? Figure 3 decomposes the cumulative change coming from the difference between exiters and incumbents (third term in equation (1)) to a part that is coming from *startup exits* – firms that exit within 5 years of entry – versus a part that is coming from *mature exits* – firms that exit after 5 years of entry. The main observation is that startup exits are the main driver of the trend in this change.

[6]Nonetheless, it is important to note that entry to the Compustat database does not automatically correspond to the economic definition of entry (firms enter Compustat when they start trading in NYSE/NASDAQ or file a report with the SEC). In particular, a different entry pattern in Compustat can be driven by a change in the timing of initial public offerings (IPOs) of private firms, which is the time when the firm is observed in the Compustat for the first time. In order to address this issue, we rely on the Worldscope database, which also records the firms' date of incorporation for a subset of the firms in Compustat. To rule out a systematic change in the timing of IPOs, we constructed the distribution of the time between incorporation and IPO for all firms and tracked key moments of this distribution over time. Our results indicate that these moments have been steady over time, and no structural change is observed. Thus, our results are not driven by structural changes in the timing of initial public offerings.

Figure 3: The COGS Share of OPEX of Exiters by Time Since Entry



Notes: This figure decomposes the cumulative change coming from exiters factor from Figure 2 to a part that is coming from startup exits versus mature exits. Startup exits are defined as firms that exit within 5 years of entry and mature exits are defined as firms that exit the dataset at least 5 years after the year of entry. Let st_{t-1} be the share of firms that exited at time $t-1$ that were startups and mt_{t-1} be the share that were mature firms. Similarly to the exercise in Figure 2, we calculate the survival rate of firms from $t-1$ to t (s_t) and define the selection among incumbents for startups as $(1-s_t) \times st_{t-1} \times (\mathbb{E}[y_{i,t-1}|\text{incumbents}] - \mathbb{E}[y_{i,t-1}|\text{startup exiters}])$. The selection among incumbents for mature firms is $(1-s_t) \times mt_{t-1} \times (\mathbb{E}[y_{i,t-1}|\text{incumbents}] - \mathbb{E}[y_{i,t-1}|\text{mature exiters}])$.

Within and Across Industry Patterns. Given the predominant role of entrants, a potential explanation for the decline is that the composition of the entering firms in terms of observables is different than the composition of incumbent firms. To analyze this possibility, we first ask: Is the decline in the aggregate COGS-to-OPEX ratio coming from a compositional effect across industries, or is it a within industry phenomenon? This can be answered with the following decomposition:

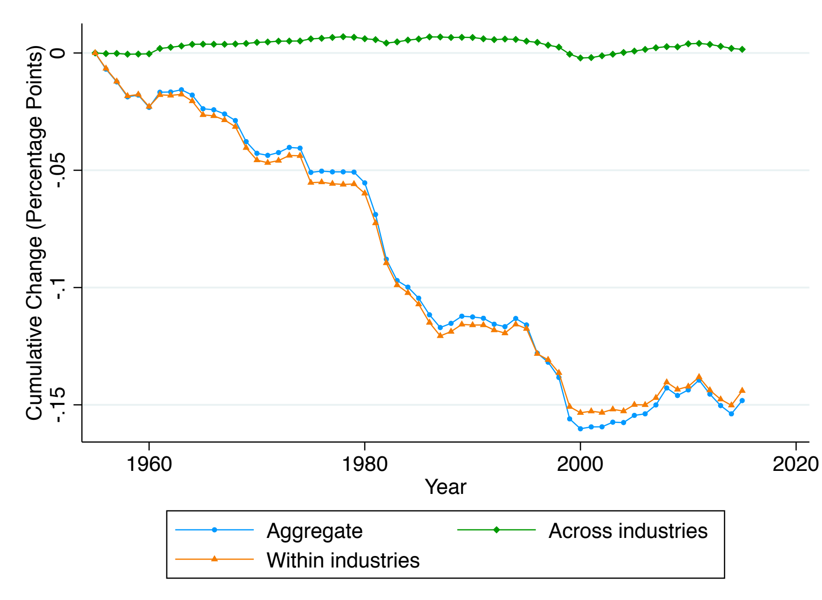
$$\Delta y_t = \underbrace{\sum_{j \in J} r_{j,t-1} \Delta y_{j,t}}_{\text{change within industries}} + \underbrace{\sum_{j \in J} y_{j,t-1} \Delta r_{j,t}}_{\text{change across industries}} + \underbrace{\sum_{j \in J} \Delta r_{j,t} \Delta y_{j,t}}_{\text{covariance term}}. \quad (2)$$

Here, y_t is the aggregate COGS-to-OPEX ratio, J is the set of industries, $y_{j,t}$ is the COGS-to-OPEX ratio within industry j and $r_{j,t}$ is the share of firms in industry j at time t. Figure 4 plots the cumulative change for each component starting in 1955. We find that almost all of the change is coming from within industries, and there is little compositional variation.[7]

Moreover, since the decline in the COGS-to-OPEX ratio is mainly driven by the extensive

[7]We have excluded the covariance term from the plot as it is negligible across the whole sample.

Figure 4: COGS share of OPEX Within and Across Industries



Notes: The figure decomposes the aggregate change in the COGS-to-OPEX ratio in the Compustat database between a change within sectors and a compositional effect across sectors. In equation (2), we have also specified a covariance term, but we omitted it from the plot because it is negligible across the whole sample. Industries are defined using the SIC 1-digit codes from the Compustat data.

margins of entry and exit, it is important to study whether this within industry pattern also holds for these margins. To investigate this question, we perform the following decomposition of selection among entrants

$$\begin{aligned}
&\underbrace{\mathbb{E}[y_{i,t}|\text{entrants}] - \mathbb{E}[y_{i,t}|\text{incumbents}]}_{\text{aggregate selection among entrants}} \qquad (3)\\
&= \underbrace{\sum_{j\in J} r_{j,t}\frac{e_{j,t}}{e_t}\left(\mathbb{E}[y_{i,t}|\text{entrants in } j] - \mathbb{E}[y_{i,t}|\text{incumbents in } j]\right)}_{\text{selection among entrants coming from within industries}}\\
&+ \underbrace{\sum_{j\in J} r_{j,t}\left(\frac{e_{j,t}}{e_t} - \frac{s_{j,t}}{s_t}\right)\mathbb{E}[y_{i,t}|\text{incumbents in } j]}_{\text{selection among entrants coming from across industries}}
\end{aligned}$$

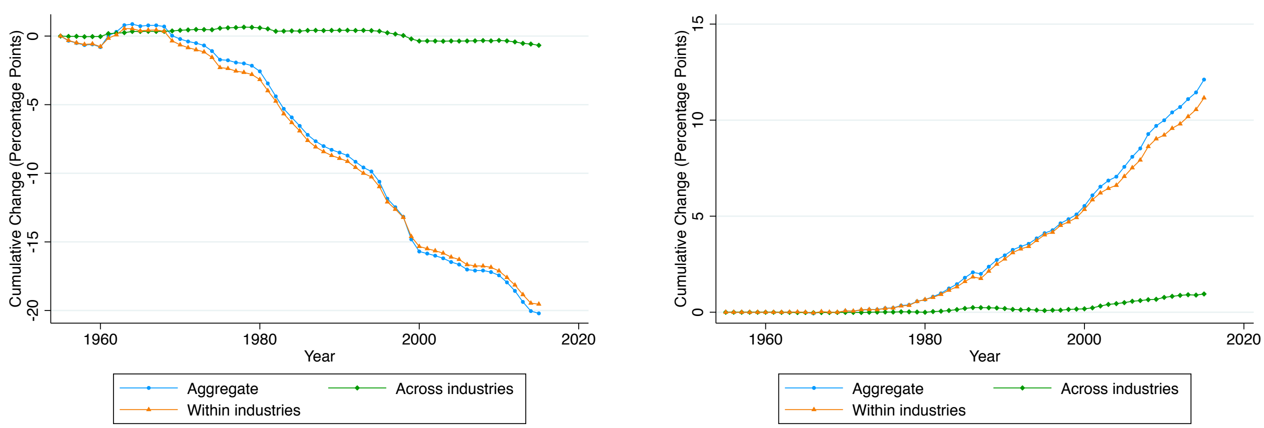
where J is the set of industries, $r_{j,t}$ is the share of firms in industry j at time t, $e_{j,t}$ is the entry rate in industry j at time t, e_t is the aggregate entry rate, $s_{j,t}$ is the survival rate in industry j at time t, and finally s_t is the aggregate survival rate. The first term in the decomposition captures the average change in the selection effect across all industries, weighing each industry by its relative rate of entry, while the second term captures whether the selection is coming from a shift of entry rates from high COGS-to-OPEX industries to low COGS-to-OPEX ratio ones. Figure 5-Panel (a) shows the two components of this decomposition. Almost all the effect is coming from selection within industries.

A similar decomposition can be done for the aggregate selection effect among exiters:

$$\underbrace{\mathbb{E}[y_{i,t-1}|\text{incumbents}] - \mathbb{E}[y_{i,t-1}|\text{exiters}]}_{\text{aggregate selection among incumbents}} \tag{4}$$
$$= \underbrace{\sum_{j \in J} r_{j,t} \frac{x_{j,t}}{x_t} \left(\mathbb{E}[y_{i,t-1}|\text{incumbents in } j] - \mathbb{E}[y_{i,t-1}|\text{exiters from } j]\right)}_{\text{selection among incumbents coming from within industries}}$$
$$+ \underbrace{\sum_{j \in J} r_{j,t} \left(\frac{s_{j,t}}{s_t} - \frac{x_{j,t}}{x_t}\right) \mathbb{E}[y_{i,t-1}|\text{incumbents in } j]}_{\text{selection among incumbents coming from across industries}},$$

where $x_{j,t}$ is the exit rate in industry j and x_t is the aggregate exit rate. Figure 5-Panel (b) shows that almost all of the effect is coming from a decline within industries. The fact that this is a within industry phenomenon does not imply that it is present in *all* industries. Below we document which industries are behind this decline in the COGS-to-OPEX ratio. This is an important step in the empirical analysis since it will be informative about potential mechanisms driving the decline.

Figure 5: Selection Effects among Entrants and Exiters: Within vs Across Industries



(a) Selection Among Entrants (b) Selection Among Exiters

Notes: This figure further decomposes the selection among entrants (panel (a)) and selection among exiters (panel (b)) for within and across industries factors. Industries are defined using the SIC 1-digit codes from the Compustat data. The "aggregate" line in panel (a) corresponds to the "entrants" factor in Figure 2, whereas the aggregate line in panel (b) corresponds to the "exiters" factor in Figure 2.

The Role of Firm's Age and Size. Previous literature found that firm-level outcomes depend on firms' age (see, e.g., Hopenhayn et al., 2018) and size (see, e.g., David et al., 2017). A natural question arises: Is the observed decline in the COGS-to-OPEX ratio the result of a change in firms' actual behavior or a compositional effect due to changes in the firms' age-size distribution? To provide an answer, we isolate the compositional effect by computing the evolution of the COGS-to-OPEX ratio after controlling for a firm's age, size, and sector. We use a firm's market share in a given year as a measure of its size and a firm's time since entry into Compustat as a measure of its age.Thus, we regress

$$y_{i,t} = \beta_y \times Year_t + \sum_s \beta_s \times Sector_{i,t} + \sum_a \beta_a \times \mathrm{Age}_{i,t} + \sum_m \beta_m \times MarketShare_{i,t} + \varepsilon_{i,t}, \quad (5)$$

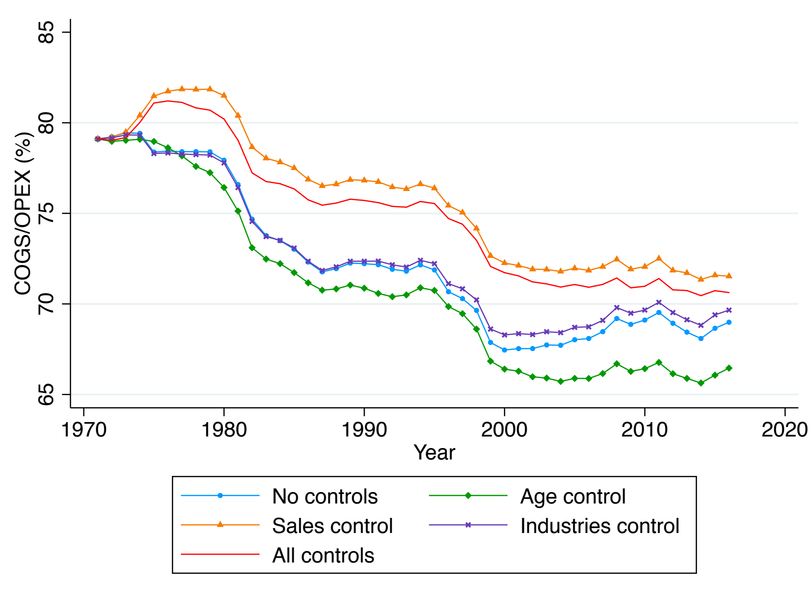
where $y_{i,t}$ is the COGS-to-OPEX ratio, *Year* denotes year fixed effects, $Sector_{i,t}$ is a categorical variable for the 1-digit SIC sector of the firm, *Age* is a categorical variable for each age group in $\{0, 1-2, 3-4, \ldots, 14-15, 16+\}$ and *MarketShare* is a categorical variable for each decile of the distribution of firms' market share.[8] We estimate several specifications, starting with the one that only includes year fixed effects and then progressively includes additional controls.

Results are reported in Figure 6. Although there exist compositional effects due to changes in the distribution of firms' size and age, we find that the inclusion of size and age controls do not eliminate the decline in the COGS-to-OPEX ratio over time. To further illustrate that this decline is due to changes in firms' actual behavior, we estimate equation (5) for each decade separately. Figure 7 plots the size profiles of the COGS-to-OPEX ratio by decade. We find that the entire size profile is declining over time. Therefore, we conclude that the observed change in firms' cost structure is not driven by changes in the distribution of firms' size and age, but instead by changes in firms' actual behavior.

Firms' Relationship with Final Consumers From the previous analysis, we conclude that the observed decline in the COGS-to-OPEX ratio is a within sector phenomenon. However, this does not imply that this ratio declined in all sectors. In fact, a broad inspection of the data shows that the decline is mostly observed in Manufacturing (SIC codes 2000-3999) and Services (SIC codes 7000-8999).A natural question is: why is it the case that the decline is only observed in those industries? Our analysis summarized in Figure 8 provides a suggestive answer: The decline in the COGS-to-OPEX ratio is more pronounced in industries that serve final consumers. We show this in two ways.

[8] Following Haltiwanger et al. (2013), we exclude the first fifteen years of the dataset for all analyses using age and we group firms older than sixteen years because we cannot know for certain a firm's IPO date for firms that were in the Compustat data since the beginning of the sample.

Figure 6: The Decline in COGS Share of OPEX after Controlling for Firms' Characteristics



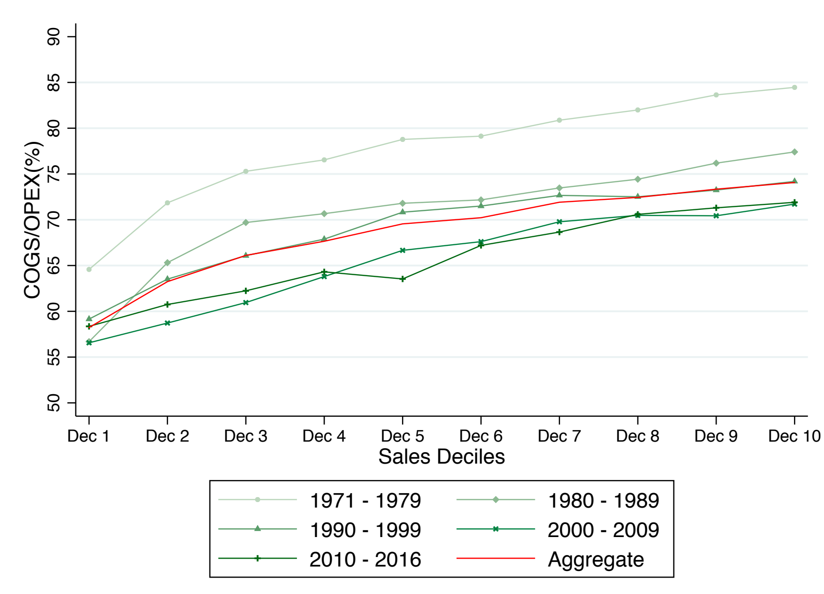
Notes: To construct the figure, we regress the firm-level COGS-to-OPEX ratio for firms in the Compustat database on time fixed effects and leave the first year (1971) as the omitted group. Then, for each year, we plot the unconditional average COGS-to-OPEX ratio for 1971 plus the corresponding year fixed effect. We estimate several specifications, starting with the one that only includes year fixed effects and then progressively include additional controls–age, sales, industries and all of them. Age is a categorical variable for the age groups in $\{0, 1-2, 3-4, \ldots, 14-15, 16+\}$. We measure a firm's market share in a given year and construct deciles of the market share variable as a control for size. Finally, we use the 1-digit SIC sector of the firm as the industries control. Following Haltiwanger et al. (2013), we exclude the first fifteen years of the dataset for all analyses using age and we group firms older than sixteen years, because we cannot know for certain a firm's IPO date for firms that were in the Compustat data since the first year.

First, using data from the Input-Output tables from BEA (*Use of Commodities by Industries*), we construct measures of "proximity" to final consumers by computing the share of an industry's output that is used for *Personal consumption expenditures*. Then, we analyze the dynamics of the COGS-to-OPEX ratio for firms sorted according to the within-industry average of this variable. Panel (a) of Figure 8 shows the evolution of the COGS-to-OPEX ratio for firms with different proximity to final consumers. Firms that operate in sectors with the lowest proximity to final consumers have a COGS-to-OPEX ratio that has been on average stagnant since the 1980s. For the remaining firms, this ratio has been further declining since then.

Second, we also separately look at industries according to their spending in advertising relative to sales (Advertising-to-Sales ratio). The goal of this exercise is to understand whether the decline in the COGS-to-OPEX ratio is more pronounced in industries that have to spend more resources in order to attract potential customers. To do this, we construct the average Advertising-to-Sales ratio by 2-digit SIC industries across the entire period.[9]

[9]We compute the average over the entire sample period to capture how business is regularly conducted in

Figure 7: Sales Profile of COGS-to-OPEX ratio over Time



Notes: The figure shows the size profiles of the COGS-to-OPEX ratio by decade. We constructed deciles for the yearly market share firms and then, for each decade separately, we regressed the COGS-to-OPEX ratio on the market share deciles, the age control, industries control and year fixed effects, where the first market share decile is the omitted group. For the figure, we plotted the unconditional COGS-to-OPEX ratio by decade for the first decile group, plus the corresponding decile coefficient.

Panel (b) of Figure 8 shows the dynamics of the COGS-to-OPEX ratio for firms in industries below and above the median of the Advertising-to-Sales ratio. The Figure shows that there has been a steady decline in the COGS-to-OPEX ratio in industries that spend a larger fraction of sales in advertising. On the other hand, for the remaining firms, the COGS-to-OPEX has been stagnant since the 1980s. We interpret these results as evidence that our main fact is more pronounced in industries that serve final consumers.

2.3 Relation to Consumer Trends

Given the importance of industries that serve final consumers in explaining the decline of the COGS-to-OPEX ratio, we analyze whether this decline is connected to recent trends in consumer behavior. In particular, the pronounced decline of the COGS share suggests that, as a result of the changing cost structure toward those not directly related to product production, such as advertising, delivery, and marketing, consumers would become less likely to switch to other products, brands, or retail chains, and vice versa.

As a first step to study the switching pattern of consumers, we analyze the shopping behavior of households across retailers over time using micro-data from ACNielsen. We find

an industry, and to avoid mechanical effects since spending in advertising is part of OPEX.

Figure 8: COGS Share of OPEX by Industry Characteristics

(a) By Use of Final Consumers

(b) By Industry Spending in Advertising

Notes: Panel (a) shows the average firm-level COGS-to-OPEX ratio for firms in the Compustat dataset, for different quartiles of year and industry specific percentage of "Personal Consumption" over "Total Use of Products" from the Input-Output Use tables for 71 industries from the Bureau of Economic Analysis (BEA). Panel (b) shows the average firm-level COGS-to-OPEX ratio for firms above and below the median in terms of industry average spending in adverting over sales in the Compustat dataset. We define industry as the SIC 2-digit codes. Both plots are sales-weighted

that over time household are buying goods from fewer retailers and having more persistent decisions about which retailer to visit. To show this, define the number of retailers each household visits per year and product group as N_{hgy}, where h denotes the household, g the product group, and y the year. Then, we aggregate this measure at the household level as:

$$N_{gy} = \sum_h w_{hgy} N_{hgy},$$

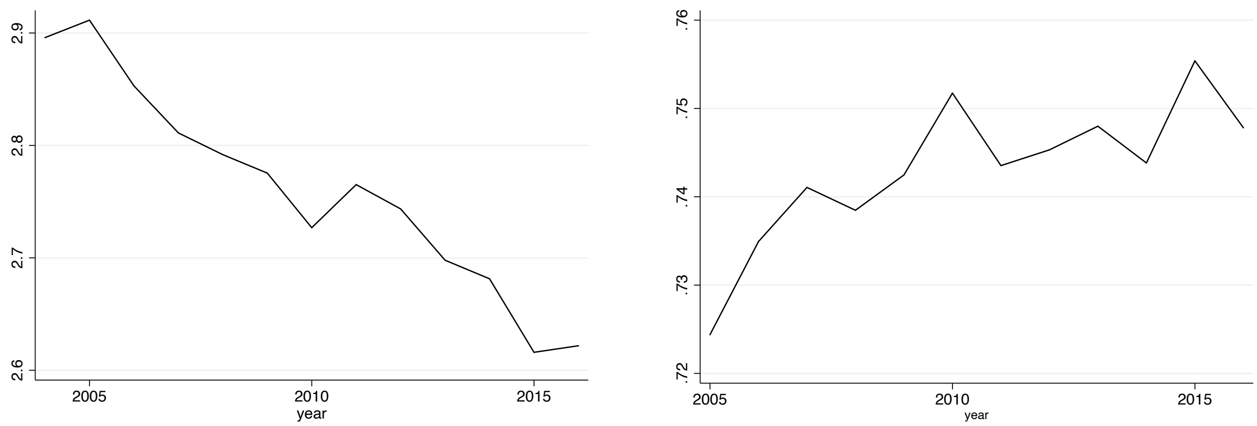
where w_{hgy} denotes the ACNielsen household sample weight (results are robust to weighting by sales, as in Neiman and Vavra (2019)). Lastly, to further aggregate N_{gy}, we take a sales-weighted average of this measure across group g:

$$N_y = \sum_g w_g N_{gy},$$

where w_g is the simple average sales share of product group g in total spending across all years in the sample (2004-2016). As shown in Panel (a) of Figure 9, the average number of retailers households visited decreased over time. Relative to the year 2004, there is about a ten percent decline in the number of retailers households visited for each product group in the year 2016.

The second fact regards the share of retailers households visited in consecutive years.

Figure 9: The Switching Pattern of Consumers across Retailers



(a) Number of Retailers Households Visited

(b) Share of Retailers Households Visited Repeatedly

Notes: The figure analyzes the households shopping pattern across retailers by using the ACNielsen Homescan Panel database. Panel (a) counts the number of retailers visited per households each year. Panel (b) measures the number of retailers households visited for the past consecutive two years over the total number of retailers households visited in a given year. In measuring the number of retailers per households, we count the retailer visits for each product group. For the aggregation, we use the household sample weight.

Define $N_{hgy,y-1}$ as the number of retailers households h visited and purchased product group g in both years $y-1$ and $y-1$. Then, we define the share of retailers revisited in two consecutive years as:

$$s_{hgy,y-1} = \frac{N_{hgy,y-1}}{N_{hgy}},$$

where $N_{hg,y}$ is the number of retailers households visited in year y. We aggregate the above measure across households and product groups following the same aggregation procedure used above. As shown in Panel (b) of Figure 9, in recent years, households tend to visit more of those retailers they visited in the previous year, rather than visiting new retailers. These results are consistent with other papers documenting long-run trends in consumer inertia. Neiman and Vavra (2019) find that households have been increasingly concentrating their spending on fewer products. Bornstein (2018) shows that consumer inertia has increased over time due to the aging of the US population. Coibion et al. (2017) document a decline in the frequency of shopping trips in the US since 1980.

So far, our analysis of consumer behavior is constrained to retail stores. As a next step, we merge the ACNielsen data with the Compustat data and connect households' purchasing behavior of products and the cost structure of these products. There are approximately 300 firms identified in Compustat that can be matched with the ACNielsen data in 2004-2016. Although the number of firms we matched is small, these firms account for approximately 22% of total sales and number of observations in the ACNielsen Homescan Panel dataset. To construct the variables of interest, let $N_{hgft,t-1}$ denote the number of products within group g and firm f purchased by household h in both time (year) $t-1$ and t. Define the probability of purchasing again the same product in the future as:

$$s_{hgft} = \frac{N_{hgft,t-1}}{N_{hgf,t-1}},$$

where $N_{hgf,t-1}$ is the total number of products within group g and firm f household h purchased at time $t-1$. In measuring these variables, we restrict the sample to only those group-household bins that are present in both $t-1$ and t.[10] Also, we exclude those firm-group bins in which all UPCs have 0 value of $N_{hgft,t-1}$.[11] Next, we aggregate the above measure across households within product group and firm by taking the weighted average:

$$s_{gft} = \sum_{h} \omega_{hgft} s_{hgft}, \tag{6}$$

where the weight ω_{hgft} is defined as:

$$\omega_{hgft} = \frac{w_{ht} E_{hgft}}{\sum_h w_{ht} E_{hgft}},$$

where E_{hgft} is the total expenditure of household h in product group g, firm f, and time t, and w_{ht} is ACNielsen's household sample weight. Figure 10 plots the share of repeated sales and the average COGS-to-OPEX ratio in Compustat data during 2004-2016. These correlation between these variables is -0.69. A low COGS-to-OPEX ratio (i.e., a high SGA-to-OPEX ratio) is associated with high repeated sales.

However, since this relationship can be affected by aggregate and industry-specific confounding facts, we analyze the relationship in a regression framework, which allows us to control for these and other firm-specific characteristics. We run the following regression analysis:

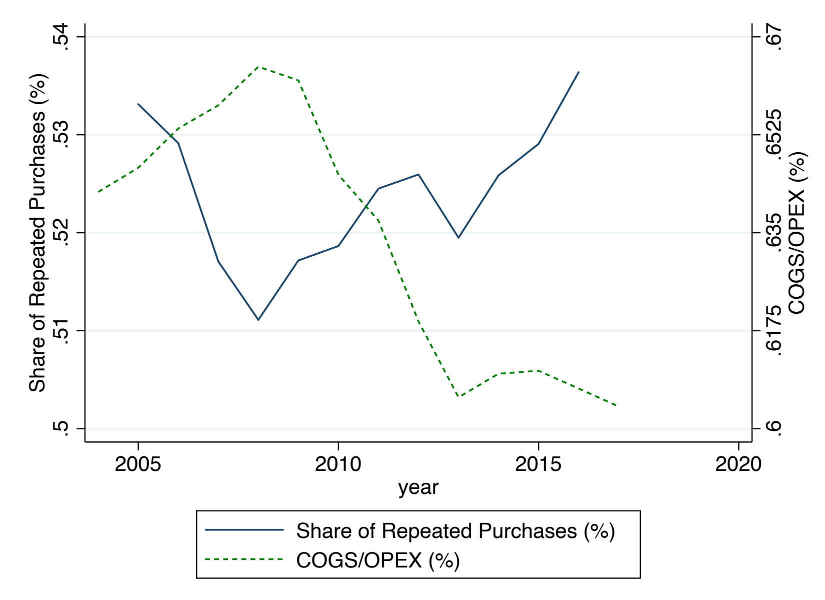
$$s_{gft} = \lambda_{gt} + \lambda_f + \beta \Delta \left(\frac{\text{SG\&A}}{\text{OPEX}} \right)_{ft} + \gamma X_{ft} + \varepsilon_{gft}, \tag{7}$$

where s_{gft} is the share of products purchased repeatedly, λ_{gt} is a group-year fixed effect, λ_f is a firm fixed effect, and X_{ft} is a vector of control variables. The coefficient of interest is β, which measured the relationship between repeated sales and a firms spending in SG&A

[10] That is, if a household never purchased a product group in two consecutive years, we exclude that product group for the household. This restriction is made to exclude those products that are not purchased frequently due to, e.g., their durability.

[11] Given a group, if a household purchased UPCs from a given firm but stopped purchasing any of the UPCs made by that firm, we exclude such firms in the analysis. Similarly, our measure does not allow for entering UPCs newly produced by firms as the denominator cannot be defined.

Figure 10: Share of Repeated Purchases and the COGS-to-OPEX Ratio



*Notes:*The share of repeated purchases is aggregated by taking a weighted average of s_{gft}, which is defined in equation (6). The weight is the average expenditure across time within firm and group. The plotted COGS-to-OPEX ratio is a simple average across firms within year using all firms in Compustat data.

relative to OPEX.

Results are reported in Table 1. The first column shows that a 10 percentage points increase in the SG&A-to-OPEX ratio is associated with a 2.86 percentage points increase in the share of repeated purchases. The following columns include variants of firm, year, product group and industry fixed effects. In our preferred specification on column (6), which includes group-year, industry-year and firm fixed effects, the share of repeated purchases increases by approximately 2.14 percentage points when the change in SG&A increase by 10 percentage point.[12] This evidence supports the interpretation that those firms directly serving final consumers raise consumer loyalty by increasing their expenses on SGA components of the total cost.

2.4 The Nature of the New Costs

What is the nature of the increasing share of SGA expenses? More importantly, is the decline in the COGS share of OPEX coming from an increase in the share of variable costs or fixed costs? The answer to this question matters for the measurement of markups and assessing the implications of the trends that motivate our study.

[12]The results are generally robust to using only data from years 2007-2015 (a period less contaminated by product entry), using NAICS 2-digit FE instead of NAICS 4-digit FE, only using manufacturers, wholesalers, and retailers, and the level of a firm's spending on SG&A instead of the SG&A-to-OPEX ratio.

Table 1: Spending in SG&A and Repeated Sales

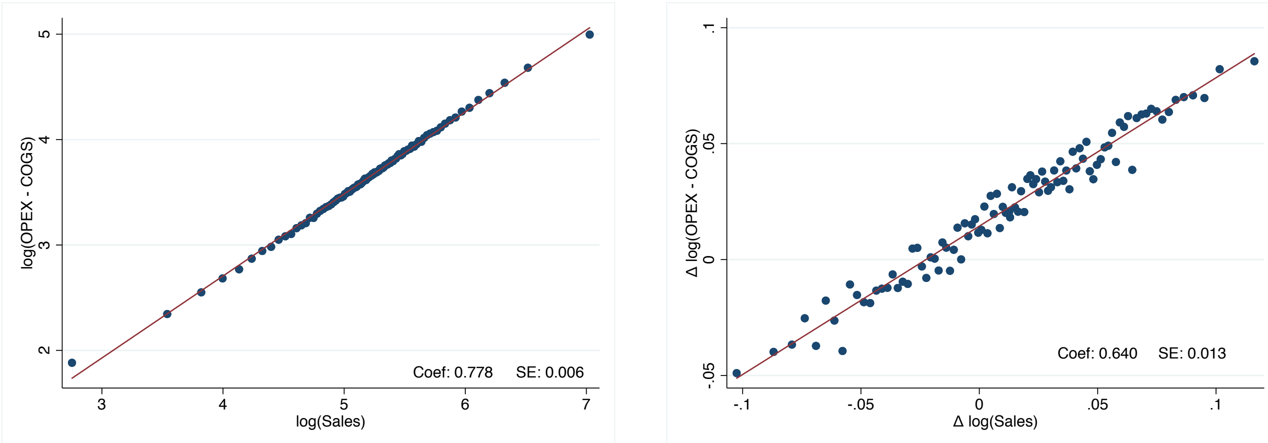
	s_{gft}: share of repeated purchases					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta\frac{\text{SG\&A}}{\text{OPEX}}$	0.286**	0.283**	0.306**	0.326**	0.371***	0.214**
	(0.131)	(0.130)	(0.134)	(0.135)	(0.094)	(0.096)
product group FE		✓	✓			
year FE		✓	✓			
NAICS 4-digit FE			✓	✓		
firm-level controls			✓	✓	✓	✓
group-year FE				✓	✓	✓
NAICS-year FE					✓	✓
firm FE						✓
Observations	18442	18441	18389	18369	18152	18124

Notes: Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; standard errors are two-way clusterd at the firm and group level. Firm-year-specific control variables include: one-year lagged log sales, log operating expenses, and assets.

Next, we provide a more detailed picture of how these costs vary with firms' production. While the current literature treats SGA expenses as fixed costs, our preliminary analysis suggests otherwise. Panel (a) of Figure 12 shows that firms' SGA expenses are almost co-linear with their sales. Panel (b) of the same figure also shows that in addition to levels, the change in SGA expenses is also strongly correlated with the change in firms' sales. These two graphs are indicative of how strongly SGA expenses vary with the level of firms' sales and point toward the fact that these expenses are more in line with a notion of a variable cost that is not a cost for production of the goods, but a cost that varies with the level of production nonetheless.

A variable nature for SGA expenses has significant implications for the current debate over the trends in market power. For instance, Traina (2019) argues that *if* SGA expenses were counted as variable costs, there is almost no trend in markups; whereas if they excluded, as in De Loecker et al. (2018), markups seem to have increased steadily over time. We aim to shed light on this debate by digging deeper into the underlying causes of these trends in the cost structure of firms and understanding how these expenses manifest themselves in different categories of costs. To this end, we will use the micro-data available in the Compustat database that categorizes firms' expenditures.

Figure 11: The Variable Nature of SGA



(a) Levels

(b) Small Changes

Notes: Panel (a) shows the bins scatterplot of yearly log of sales and yearly log of OPEX minus COGS for firms in the Compustat dataset. In the plot, we group the x-axis variable (log of sales) into 100 equal-sized bins and calculate the average of the log of sales and the log of OPEX minus COGS (y-axis) within each bin. Each one of these 100 points is plotted in panel (a) as a scatterplot. We also plot the fit line based on the two regressions. Both the averages for the scatterplot and the fit line are controlled for year and firm fixed effects. Industries are defined as SIC 2-digit codes. For panel (b) we do a very similar exercise, but this time we plot the bins scatterplot for the difference in the log of OPEX minus COGS for time t and time $t-1$ (y axis) and the difference of log of sales for time t and time $t-1$. For panel (b) we restrict the sample to observations where the change in the log of sales was between -0.1 and +0.1. Once again, we control for year and firm fixed effects, where industries are defined as SIC 2-digit codes. The coefficients and the standard error for both plots represent the results for the best fit line, with standard errors clustered on the firm level.

3 Model

Our empirical results point towards two separate channels that could potentially have different macroeconomic implications. On the one hand, entrants seem to spend more on customer acquisition in recent years, which is driven by a lower propensity of consumers to switch away from incumbents. Since these costs do not fit the definition of cost of goods sold, they are not registered in COGS and eventually contribute to a higher *measured* markup, since they affect the price of the good but are not considered as a part of the firms' cost for production. That said, these costs are not necessarily inefficient. If an entrant is highly productive in producing their variety but their product is not purchased due to search frictions, the social planner would allocate some expenses toward customer acquisition for such a firm.

On the other hand, the lower propensity of switching for consumers gives the incumbents higher market power and allows them to charge higher markups to begin with. This channel also contributes to higher measured markups, but in an inefficient way.

Thus, the trends that we observe in the data for the fall in the COGS share of OPEX and lower intensity of search on the side of households seem to be an amalgamation of two phenomena that go hand in hand: if entrants have to spend more for customer acquisition, it has to be the case that incumbents have higher market power. The main goal of our model is to provide us with a framework that allows us to decompose the strength of each of these two channels. From this perspective, the model needs to have two main ingredients: (1) endogenous firms dynamics to create a realistic distribution of size among incumbents in order to capture the relationship between COGS-to-OPEX ratio and sales, and (2) endogenous consumer taste for products in order to capture the concept of costumer acquisition for firms.

Our second objective in writing a model is to examine whether the SGA expenses are partly fixed costs or variables costs (or a combination of both). The distinction is important because if a part of SGA expenses are variable, then there is a wedge between the marginal cost of production and the marginal cost of a sale. While the former captures only the cost of producing an additional unit, the latter also takes into account the cost of getting that additional unit to the hands of the consumer. If such a wedge exists, then it also manifests itself in measuring the markups. In particular, if it is the variable cost component of the SGA expenses that are increasing over time, it is potentially the case that the price over marginal cost of production markups are rising, but the price over marginal cost of sales markups are not. From this perspective, a third main ingredient for the model is to allow for both fixed and variable SGA expenses.

3.1 Modeling Environment

Time is discrete and is indexed by $t \in \{0, 1, 2, \dots\}$. The economy consists of a representative household and a measure of firms that produce weakly substitutable goods for the household. Firms and their varieties are indexed by $i \in [0, \bar{N}_t]$ where $\bar{N}_t$ is the measure of operating firms at time t.

Households. At any given time, the representative household supplies labor to the firms in a competitive labor market and has preference for the existing varieties at that period. We assume that preferences at this stage are governed by deep habits as in Ravn et al. (2006).

Formally, the household solves:

$$\max \sum_{t=0}^{\infty} \beta^t \left[\frac{U_t^{1-\gamma}}{1-\gamma} - \xi \frac{L_t^{1+\psi}}{1+\psi} \right] \quad \text{(household's problem)}$$

$$s.t.\ U_t = \left[\int_0^{\bar{N}_t} m_{i,t}^{\frac{1}{\sigma}} \left(c_{i,t} - \delta \bar{c}_{i,t-1} \left(\frac{c_{i,t}^*}{U_t} \right)^{\frac{\eta}{\sigma}} \right)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}} \quad \text{(preference aggregator)}$$

$$\int_0^{\bar{N}_t} p_{i,t} c_{i,t} di \leq W_t L_t + \int_0^{\bar{N}_t} \Pi_{i,t} di \quad \text{(budget constraint)}$$

where $m_{i,t}$ is the taste of the consumer for variety i at time t, $\bar{c}_{i,t-1}$ is their (external) habit that is given by $\bar{c}_{i,t-1} = c_{i,t-1}$, and $c_{i,t}^*$ is the counterfactual demand of a household who does not have external habits. When $\eta = 0$, the preferences are identical to Ravn et al. (2006).

It follows that the household's labor supply curve is given by the standard intratemporal Euler equation

$$\xi L_t^{\psi} = W_t U_t^{-\gamma} \quad \text{(labor supply)}$$

Moreover, demand for every variety is given by

$$c_{i,t} = \delta c_{i,t-1} \left(\frac{p_{i,t}}{P_t} \right)^{-\eta} + m_{i,t} U_t \left(\frac{p_{i,t}}{P_t} \right)^{-\sigma} \quad \text{(demand for varieties)}$$

where $P_t \equiv [\int_0^{\bar{N}_t} m_{i,t} p_{i,t}^{1-\sigma} di]^{\frac{1}{1-\sigma}}$ is the aggregate price index. Note that this demand function implies an elasticity between the two bounds η and σ. In fact, if we define the share of "captive" demand relative to total demand as

$$s_{i,t} \equiv \frac{\delta c_{i,t-1} \left(p_{i,t}/P_t \right)^{-\eta}}{\delta c_{i,t-1} \left(p_{i,t}/P_t \right)^{-\eta} + m_{i,t} U_t \left(p_{i,t}/P_t \right)^{-\sigma}} \quad \text{(share of captive demand)}$$

we can write the elasticity of demand as a weighted average of η and σ

$$\varepsilon_{i,t} \equiv -\frac{\partial c_{i,t}}{\partial p_{i,t}} \frac{p_{i,t}}{c_{i,t}} = (1 - s_{i,t})\sigma + s_{i,t}\eta \quad \text{(elasticity of demand)}$$

Henceforth, we assume $1 < \eta < \sigma$ to (1) ensure that demand for all goods is elastic and (2) focus on the case where demand that is driven by habits (captive demand) is less elastic than demand driven by new taste.

Firms. At each period, a measure ρ of potential entrants are born, upon which they draw a life-time productivity scale of $z \in \mathbb{R}_+$ from a distribution $\Gamma_z(.)$. After drawing their productivity, potential firms can decide to enter the economy in which they will face demand for their variety but will be subject to an overhead cost of $\chi > 0$ every period in terms of units of labor. If a potential entrant decides to enter, they become an incumbent.

Incumbent firms take aggregate prices and their demand from the household side as given and choose the price for their product along with their demand for labor. These firms can hire labor from a competitive labor market at wage W_t to either produce their variety with production function $y_{i,t} = z_i l_{i,p,t}^{\alpha}$ (with $\alpha < 1$) or to invest in demand by creating new matches through $m_{i,t} = \bar{m} l_{i,s,t}^{\phi}$ (with $\phi < 1$), where $m_{i,t}$ is the taste of the consumer for product i. Therefore, we assume that firms can directly affect the consumers' taste or their product independent of price. We interpret this as investing in demand shifters such as advertising, marketing, or mitigating search frictions that allows the firm to reach a broader set of potential consumers. Finally, we assume that incumbent firms exit at an exogenous rate of $1 - \nu$.

Incumbents' Problem. Formally, an incumbent firm's at a given time t is to solve

$$
\begin{aligned}
v_t(y_{i,t-1}, z_i) \equiv & \max_{\{p_{i,t+h}, l_{i,s,t+h}, l_{i,p,t+h}\}_{h\geq 0}} \sum_{h=0}^{\infty} (\beta\nu)^h \left[p_{i,t+h} c_{i,t+h} - W_t(l_{i,p,t+h} + l_{i,s,t+h} + \chi)\right] \\
s.t. \quad & c_{i,t+h} = \delta c_{i,t+h-1} \left(\frac{p_{i,t+h}}{P_{t+h}}\right)^{-\eta} + U_{t+h} \bar{m} l_{i,s,t+h}^{\phi} \left(\frac{p_{i,t+h}}{P_{t+h}}\right)^{-\sigma}, \quad y_{i,t-1} \geq 0 \quad \text{given.} \\
& y_{i,t+h} = z_i l_{i,p,t+h} \geq c_{i,t+h} \qquad \text{(incumbents' problem)}
\end{aligned}
$$

Entrants' Problem. A potential entrant at a given time t with productivity scale z_i would enter only if their lifetime value is positive:

$$
\max_{\{\text{enter,not enter}\}} \{v_t(0, z_i), 0\} \qquad \text{(entrants' problem)}
$$

Definition of Equilibrium. An equilibrium for this economy is an allocation for households $\{(c_{i,t}, U_t, L_t)\}_{t\geq 0}$, a set of prices and labor demands for incumbent firms $\{(p_{i,t}, l_{i,p,t}, l_{i,s,t}) : i \text{ is incumbent at } t\}_{t\geq 0}$, an entry decision for all potential entrants $\{(1_{i,enter}) : i \text{ is entrant at } t\}_{t\geq 0}$, and a set of aggregate prices $\{W_t, N_t, P_t\}_{t\geq 0}$ such that

1. given prices, household's allocation solves her problem.

2. given aggregate prices and household's demand for their varieties, incumbent firms' allocation solves their problem.

3. given aggregate prices, the entrants' allocation solves their problem.

4. market clears:

$$c_{i,t} = y_{i,t}, \forall i \in [0, \bar{N}_t], \forall t \geq 0$$
$$L_t = \int_0^{\bar{N}_t} (l_{i,p,t} + l_{i,s,t} + \chi) di \forall t \geq 0$$

Moreover, we call an equilibrium to be *stationary* if $\{W_t, U_t, P_t\}_{t \geq 0}$ are constant over time.

3.2 Characterization

Since all prices matter in relative terms, we normalize the aggregate price $P_t = 1$, and start by focusing on the stationary equilibrium. Let $\{W, U\}$ denote the the wage and the utility of the household in this equilibrium. Then, the problem of an incumbent firm with productivity scale z and lagged demand y_{-1} is characterized by

$$v(y_{-1}, z) = \max_{p, l_s} \{py - W(\frac{y}{z})^{\alpha^{-1}} - Wl_s - W\chi + \beta\nu v(y, z)\} \quad \text{(incumbents value function)}$$
$$s.t. \quad y = \delta y_{-1} p^{-\eta} + U\bar{m} l_s^{\phi} p^{-\sigma}$$

Moreover, a potential entrant with productivity scale z will enter in this stationary equilibrium if and only if

$$v(0, z) \geq 0 \quad \text{(entrants' decision)}$$

Two notions of marginal costs. Contrary to classic models of firm production where the only costs for firms are costs of production, our model creates a wedge between the cost of *producing* a good, and the cost of *selling* a good. While the former only takes into account the cost of production labor, the latter requires generating enough matches in order to sell the goods that are produced.

Furthermore, not all costs that are related to selling goods are fixed. In addition to the overhead cost that firms have to pay, they can optimally choose to which scale they want to create new matches. Therefore, the cost of sales are also partially variable, and contribute to the firms' marginal costs. Hence, the wedge between these two types of costs also creates a wedge between the marginal cost of production, and the marginal cost of a sale. Since our analysis of markups are going to heavily rely on these two types of marginal costs, we formally define these notions in our framework.

Definition 1. For a given price p, let l_p^* and l_s^* be the cost minimizing labor demands for the firm at that price in order to produce and *sell* at least $\bar{y}$ units of the their variety at that price. Then, the marginal costs of production (mc_p) and a sale (mc_s) at $\bar{y}$ are formally defined as

$$mc_p = \frac{\partial W l_p^*}{\partial \bar{y}} \qquad \text{(marginal cost of production)}$$

$$mc_s = \frac{\partial W(l_p^* + l_s^*)}{\partial \bar{y}} \qquad \text{(marginal cost of a sale)}$$

3.3 A Special Case with Analytical Results

In this section we consider the special case of a model in which $\delta = 0$. In addition we assume constant returns to scale in production ($\alpha = 1$), and that the distribution of productivity for potential entrants, $\Gamma_z(.)$, is a Pareto distribution with the tail parameter $\zeta > \max\{1, \frac{\sigma-1}{1-\phi}\}$.

In this case, there is no captive demand in the firms' problem and an incumbent firm's problem boils down down to

$$v(z) = \max_{\{p,l_s\}}\{py - Wz^{-1}y - Wl_s - W\chi + \beta\nu v(z)\}$$
$$s.t.\ y = U\bar{m}l_s^{\phi}p^{-\sigma}$$

Markups. The constant returns to scale assumption implies that a firm's marginal cost is simply the ratio of wage over their productivity. It then follows directly form the incumbent firms' problem that price over marginal cost of production markups have the classic representation of

$$\mu_p \equiv \frac{p}{Wz^{-1}} = \frac{\sigma}{\sigma - 1} \qquad \text{(production markup)}$$

In spite of its conventional representation, the production markup is no longer the correct measure of the markup is this economy as it ignores the cost of selling of a marginal unit that comes from the variable cost nature of generating new matches. In fact, we can analytically derive and characterize this gap in this simple case:

Proposition 1. *Let μ_s denote the price markup of the firm over the marginal cost of a sale. Then,*

1. *μ_s is given by*

$$\mu_s = \frac{p}{mc_s} = \frac{\sigma}{\sigma - 1 + \phi} \in [1, \frac{\sigma}{\sigma - 1}] \qquad \text{(sales markup)}$$

which increases with the elasticity of demand σ and decreases with the elasticity of the matching function with respect to labor allocated to sales, ϕ.

2. *the wedge between the production and sales markups is given by*

$$\tilde{\mu} \equiv \frac{\mu_p}{\mu_s} = 1 + \frac{\phi}{\sigma - 1} \in [1, \infty) \qquad \text{(markup wedge)}$$

which increases with ϕ and decreases with σ.

The first part of the proposition shows that the production markup is always larger than the sales markup. This is expected since the marginal cost of sales is defined to take the marginal cost of production into account. The more interesting part of this result, however, is that the inequality is strict if the elasticity of the matching function with respect to labor is strictly positive, $\phi > 0$.

The second part of the proposition is at the crux of our argument in this paper. It shows that wedge between the two markups increases with the elasticity of matching with respect to labor, but more importantly, it shows that this wedge is larger at lower levels of elasticity of demand.

When demand is less elastic, consumers are more attached to the incumbent firms, which implies that firms have to spend more resources to generate new matches. Nonetheless, since this is not a cost associated with production, it does not manifest itself in the marginal cost of production and just shows itself in marginal cost of sales, which expands the wedge between the two markups.

This force is going to be at the heart of our quantitative analysis. In fact, letting the data select the elasticity of matches to labor and elasticity of demand, we are interested in measuring this wedge, and how it has grown over time. If the rise in consumer inertia has dragged down the elasticity of demand in the past few decades, it would partially explain a rise in production markups. However, the core message of this simple model is that the rise in production markups is a gross *over-estimation* of the the rise in sales markups, as long as the data supports that $\phi > 0$.

The Relationship between SGA and Sales. So far, we have established that including a variable cost component in generating sales creates a wedge between the production and sales markups. In this section, we tie this back to our analysis of the the correlation between sales and SGA. In particular, we are interested in analyzing what this correlation identifies within the context of the model.

To this end, we define the SGA expenses of a firm as the sum of their overhead costs and cost of labor associated with generating new matches:

$$SGA_{i,t} = W\chi + Wl_{i,s,t} \tag{8}$$

Moreover, we define the sales of a firm as

$$Sales_{i,t} = p_{i,t}y_{i,t} \tag{9}$$

Proposition 2. *Consider the regression:*

$$SGA_{i,t} = Constant + \beta_s \times Sales_{i,t} \tag{10}$$

Then $\beta_s = \frac{\phi}{\sigma}$ *identifies the ratio of the elasticity of generating new matches with respect to labor to elasticity of demand.*

4 Quantitative Analysis

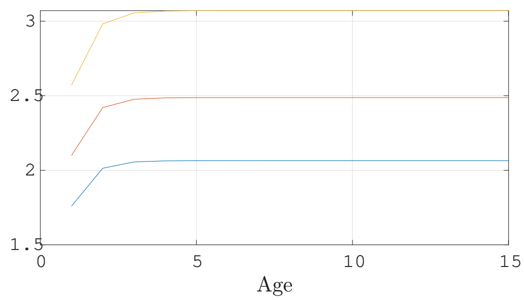
4.1 Calibration Strategy

Fixed parameters

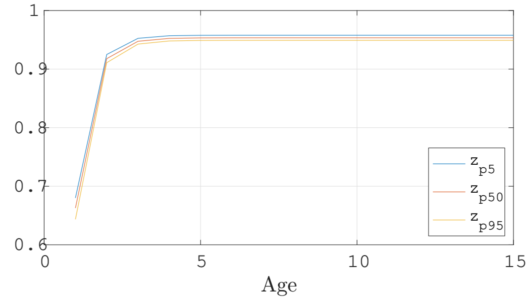
Calibrated parameters

Model Fit

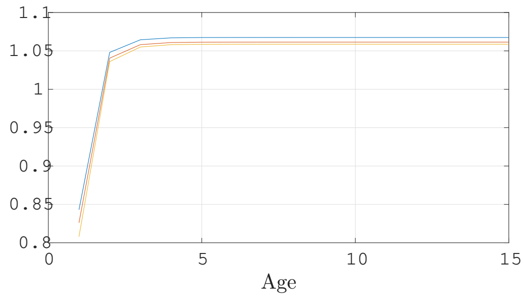
Figure 12: The Variable Nature of SGA



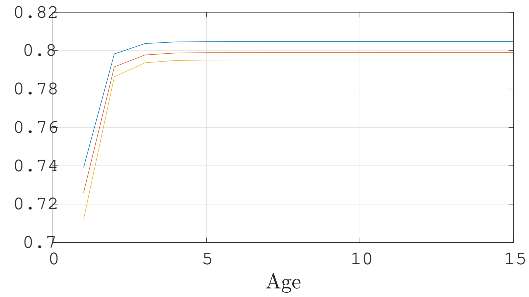
(a) Output



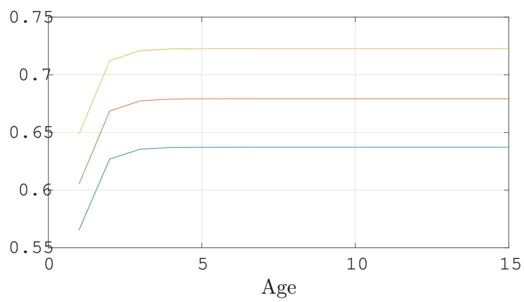
(b) Price



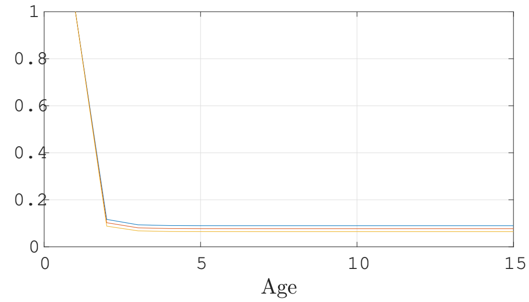
(c) Markup (prod.)



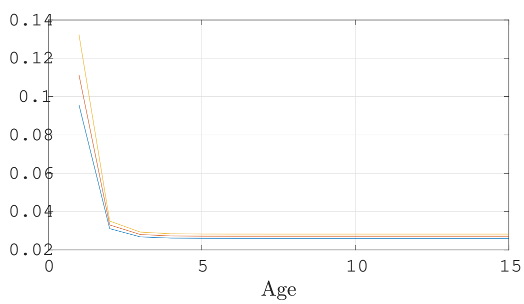
(d) Markup



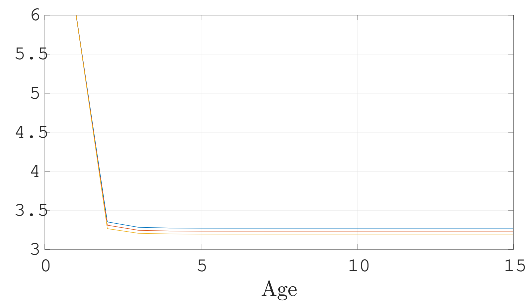
(e) COGS/OPEX



(f) Share new customers



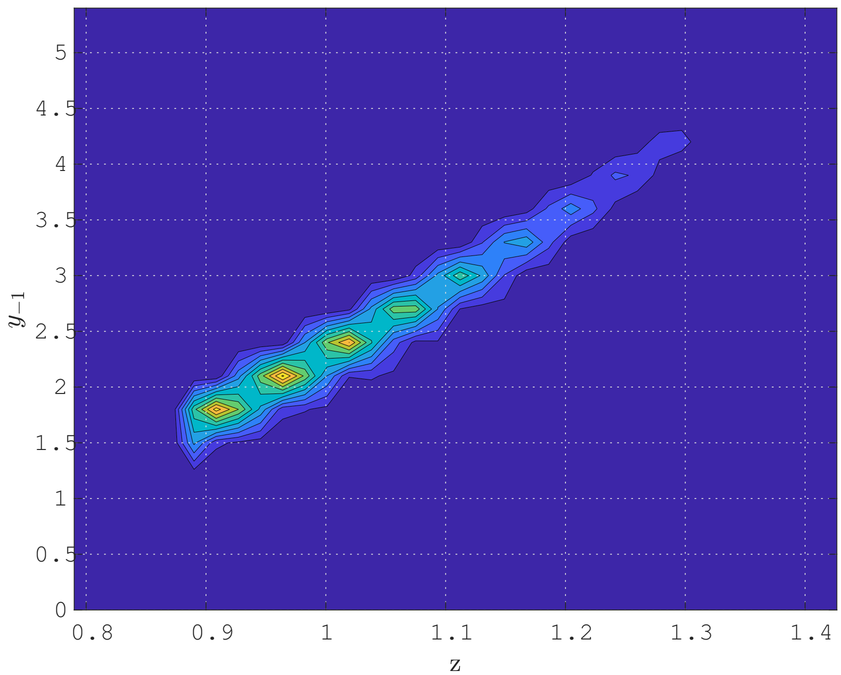
(g) l_s



(h) Elasticity

Notes:

Figure 13: Steady-State Distribution of Firms



Notes: This figure plots the distribution of firms in steady state by a firm's productivity and measure of captive consumption.

4.2 Firms' Cost Structure in Steady State

4.3 Implications for Measurement of Markups

5 The Decline in the COGS-to-OPEX Ratio

5.1 The Role of Habits

5.2 Implications for the Rise in Markups

5.3 Alternative Mechanisms

6 Conclusion

[TBC]

References

Bornstein, Gideon, "Entry and Profits in an Aging Economy: The Role of Consumer Inertia," Technical Report 2018. Mimeo.

Coibion, Olivier, Yuriy Gorodnichenko, and Dmitri Koustas, "Consumption inequality and the frequency of purchases," 2017.

David, H, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The fall of the labor share and the rise of superstar firms," 2017. CEPR Discussion Paper No. DP12041.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How costly are markups?," Technical Report, National Bureau of Economic Research 2018.

Gouin-Bonenfant, Emilien, "Productivity dispersion, between-firm competition and the labor share," 2018. Mimeo.

Gutierrez, German, Callum Jones, and Thomas Philippon, "Entry Costs and the Macroeconomy," 2019. Mimeo.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hopenhayn, Hugo, Julian Neira, and Rish Singhania, "The Rise and Fall of Labor Force Growth: Implications for Firm Demographics and Aggregate Trends," 2018. Mimeo.

Kehrig, Matthias and Nicolas Vincent, "The micro-level anatomy of the labor share decline," Technical Report, National Bureau of Economic Research 2018.

Loecker, J De, J Eeckhout, and G Unger, "The Rise of Market Power and the Macroeconomic," 2018. Mimeo.

Neiman, Brent and Joseph S Vavra, "The Rise of Niche Consumption," Technical Report, National Bureau of Economic Research 2019.

Ravn, Morten, Stephanie Schmitt-Grohé, and Martin Uribe, "Deep habits," *The Review of Economic Studies*, 2006, *73* (1), 195–218.

Traina, James, "Is aggregate market power increasing? production trends using financial statements," 2019. Mimeo.

Weiss, Joshua, "Intangible Investment and Market Concentration," 2019. Mimeo.

Zhang, Lichen, "Intangibles, Concentration and Labor Share," 2019. Mimeo.

APPENDIX FOR ONLINE PUBLICATION

A Data construction

Herein we describe the cleaning process to generate the panel dataset used for the paper's analysis. Our cleaning process was guided and is very similar to the previous literature that used Compustat to analyze the cost of production over time (see for instance De Loecker et al. (2018); Traina (2019)).

A.1 Variables

We download and construct the following variables from Compustat:

- *Global company key* (mnemonic `gvkey`): Compustat's firm id.
- *Cusip*: identifier for "a specific security issue of a company."
- *Year* (mnemonic `fyear`): the fiscal year.
- *ISO country code, incorporation* (mnemonic `FIC`): indicates the country in which a company was incorporated.
- *Selling, general and administrative expense* (mnemonic `XSGA`): the SG&A sums "all commercial expenses of operation (such as, expenses not directly related to product production) incurred in the regular course of business pertaining to the securing of operating income."
- *Costs of goods sold* (mnemonic `COGS`): the COGS sums all "expenses that are directly related to the cost of merchandise purchased or the cost of goods manufactured that are withdrawn from finished goods inventory and sold to customers."
- *Operating expenses, total* (mnemonic `XOPR`): OPEX represents the sum of COGS, SG&A and other operating expenses.
- *Sales (net)* (mnemonic `SALE`): this variable represents gross sales, for which "cash discounts, trade discounts, and returned sales and allowances for which credit is given to customer" are discounted from the final value.
- *Assets, total* (mnemonic `AT`).
- *Acquisitions* (mnemonic `AQC`): this variable constitutes the costs relating to aqcuisition of another firm.
- *Capital*: for the first observation of each firm, we set capital to be equal to the gross property, plant and equipment value (mnemonic `PPEGT`). Then for susbsequent years we use net PPE (mnemonic `PPENT`) and add the difference from $\texttt{PPENT}_t$ and $\texttt{PPENT}_{t-1}$. We deflate the difference in `PPENT` by the investment goods deflator from NIPA's nonresidential fixed investment good deflator (line 9).

- *Advertising* (mnemonic XAD): this variable contains the cost of media advertising and promotional expenses.
- *North american industry classification code* (mnemonic NAICS): the NAICS classify firms by economic activities and was adopted by the Office of Management and Budget (OMB) in 1997.
- *North american industry classification system, historical* (mnemonic NAICSH): the historical NAICS code tracks a company's industry over time.
- *Standard industry classification code* (mnemonic SIC): the SIC is a four-digit classification of a company's operations.
- *Standard industrial classification, historical* (mnemonic SICH): the historical SIC code tracks a company's industry over time.
- *Company's initial public offering date* (mnemonic IPODATE).
- *Age*: given that the initial public offering date was missing for a large portion of our dataset, we calculated age as the fiscal year of a given observation minus the first year that we observe a firm in the dataset. According to Compustat, a firm enters the dataset after it starts providing consistent accessible annual reports trading on a U.S. exchange market, i.e. after its IPO. Following Haltiwanger et al. (2013), we exclude the first fifteen years of the dataset for all analyses using age and we group firms older than sixteen years, because we cannot know for certain a firm's IPO date for firms that were in the Compustat data since the first year.

We used NIPA Table 1.1.9. GDP deflator (line 1) to generate the real value for the variables sale, COGS, XOPR, XSGA, XAD and AT.

Next, we downloaded data from the Bureau of Economic Analysis's (BEA) input output accounts data. We downloaded 71-industries use tables from 1997 to 2018 and utilized the following variables:

- *IOCode*: BEA's id for each industry.
- *Total use of products*: total amount consumed from each industry.
- *Personal consumption expenditures*: total amount consumed for personal consumption by each industry.
- *Share of consumer consumption*: for each industry we define this variable as the ration of the total personal consumption expenditures to the total use of products.

A.2 Data Cleaning

Sample Selection

We downloaded the dataset "Compustat Annual Updates: Fundamentals Annual," from Wharton Research Data Services, from Jan 1950 to Dec 2016. The following options were chosen:

- Consolidated level: C (consolidated)
- Industry format: INDL (industrial)
- Data format: STD (standardized)
- Population source: D (domestic)
- Currency: USD
- Company statyus: active and inactive

The raw dataset contains 436,891 observations for 33,327 firms.
Next, we took the following steps for the cleaning process:

1. To select American companies, we filtered the dataset for companies with Foreign Incorporation Code (FIC) equal to "USA."
2. We replace industry variables (`sic` and `naics`) by their historical values whenever the historical value isn't missing.
3. We drop utilities (`sic` value in the range [4900, 4999]) because their prices are very regulated and financials (`sic` value in the range [6000,6999]) because their balance sheets are exceptionally different than the other firms in the analysis.
4. To ensure quality of the data, we drop missing or non-positive observations for sales, assets, COGS, OPEX and capital. We also exclude observations in which acquisitions are more than 5 % of the total assets of a firm.
5. A portion of the data missing for sales, COGS, OPEX, and capital in between years for firms. We input these values using a linear interpolation, but we do not interpolate for gaps longer than two years. This exercise inputs data for 5.9% of our sample.
6. Our next step is to trim the data for outliers on the firm level. We tag firms according to the following four groups and drop them if they are in any of the groups. This exercise drops 534 firms and 9,118 observations of the sample:
   - group 1: the firm's maximum real sales is in the top 0.5 percent of the distribution or the firm's minimum real sales is in the bottom 0.5 percent of the distribution.
   - group 2: the firm's maximum real COGS is in the top 0.5 percent of the distribution or the firm's minimum real COGS is in the bottom 0.5 percent of the distribution.

- group 3: the firm's maximum real OPEX is in the top 0.5 percent of the distribution or the firm's minimum real OPEX is in the bottom 0.5 percent of the distribution.
- group 4: the max COGS/OPEX ratio change in absolute value of a firm is larger than 100%.

Our final dataset contains 241,204 observations for 20,034 firms.

Merging Datasets:

Once the Compustat data was cleaned, we merged it to the BEA input-output data and the WorldScope by taking the following steps, in order:

1. We merged NAICS codes to each of the IOCode industries, as provided by the BEA.
2. We improved the number of missing NAICS codes in Compustat by using a concordance table from the SIC industry codes to NAICS (our data is not missing SIC codes).
3. We merged the input-output data to Compustat using the NAICS code in a best merge case scenario, i.e. we first attempt to merge it on the 6-digit industries, for those that fail to merge we attempt the merge on the 5-digit industries, etc.
4. Finally, we merged the Compustat to WorldScope using the cusip id.