Advertising, Brand Preferences and Market Structure

Lun Li*

September 23, 2020

Abstract

Over the past decade, U.S. aggregate markup and concentration have been steadily declining. This paper explores how the evolution of of consumer brand preferences - positive demand shifters for products from certain brands - induced by innovations in advertising technologies, contribute to these changes in the market structures for a wide range of consumer goods. First, I document empirical evidence for changes in U.S. market structures and advertising technologies from 2010 to 2016, using a unique firm-level panel data set that I construct. Second, I develop a multi-product, multi-sector heterogeneous firm model with endogenous advertising and entry/exit decisions. I then calibrate the model with empirical estimates from the micro-data, and show that advertising firms have higher brand preferences than non-advertising ones. Finally, counterfactual analysis suggests that if the advertising cost structure has not changed during the sample period, U.S. aggregate markup and Herfindahl would both *increase*.

*I am indebted to Greg Kaplan, Joseph Vavra and Chad Syverson for their continuous guidance and support. I also thank Brad Shapiro, Fernando Alvarez, Robert Shimer, Harold Uhlig, Simon Mongey and participants at the University of Chicago Capital Theory working group for discussion and comments. Researcher(s) 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. I also gratefully acknowledge the financial support from Becker Friedman Institute for the purchase of GS1 US data. This work was completed in part with resources provided by the University of Chicago Research Computing Center. All errors are my own. Email: lunl@uchicago.edu.

1 Introduction

Over the past decade, the aggregate market structure in the U.S. has changed significantly. Using supermarket scanner data from 2010 to 2016, I show that the average markup and concentration for consumer packaged goods have declined by 5% and 10% respectively. During the same period, the revenue of online advertising has soared from $26 billion to $73 billion in the U.S. The proliferation of internet advertisements creates additional channels for firms to promote their products, and challenges the pricing model of the traditional advertising industry. This paper explores the role of advertising innovation as a driver to the transformation of aggregate market structure over time.

My main hypothesis is that firms use advertisements to compete for higher levels of brand preferences among shoppers. By "brand preferences", I mean the tendency for consumers to purchase certain brands over others despite of similar prices or qualities. The existence of brand preferences is well-documented in the marketing literature (Bronnerberg, Dube and Gentzkow 2012), but the term has not been widely adopted by macro or IO economists. This paper provides a quantitative framework to estimate the magnitude of brand preferences at the firm level, and also to evaluate the importance of advertising as a source of firm heterogeneity in sales and market shares.

The economic effects of advertising have intrigued generations of economists, dating back to Marshall (1890, 1919) and Chamberlin (1933). Broadly speaking, the prevalent theories on the role of advertising fall into one of three categories: to persuasive consumers into altering their tastes (Braithwaite 1928, Kaldor 1950); to convey information (Ozga 1960, Stigler 1961); or to complement actual consumption (Stigler and Becker 1977). Starting from the 1960s and 1970s, empirical works that study the relationship between advertising and other economic variables begin to emerge[1]. But as Bagwell (2007) points out, many of the early empirical studies are vulnerable to the endogeneity problem where firms with higher sales usually spend more on advertising. In the past decade, developments in online advertising have allowed empirical studies to use large-scale randomized controlled trials (RCTs) to

[1]See Comanor and Wilson (1979) for a survey of empirical studies on advertising in the 1960s and 1970s.

identify the causal relationship between advertising and sales. But these studies commonly find little or no evidence of any measurable benefit from advertising (Blake, Noskos and Tadelis 2014, Lewis and Rao 2015, Shapiro, Hitsch and Tuchman 2020).

This paper differs from the previous literature in the following three ways. First, instead of relying on reduced-form regressions, this paper develops a structural model with nested demand systems and multi-product heterogeneous firms, following the monopolistic competition framework introduced by Hottman, Redding and Weinstein (2016). In addition to the the original model, in this paper I assume firms can endogenously decide whether to advertise, and if so choose the amount to spend on advertisements. The rich structures of this model make it possible to study firm's optimal decision rule for advertising spending, and to decompose the general equilibrium effects of an unexpected change to advertising costs. Second, this paper uses a novel firm-level panel data set with advertising spending and product sales information. Our sample is created by fuzzy-matching firm names between supermarket scanner datasets and Nielsen Ad Intel, a large-scale database that contains spend and impression information for each ad covered from 2010 to 2016. The merged data includes more than 3000 advertising firms across over 400 narrowly defined product categories, and is arguably the most comprehensive dataset ever studied for this purpose. Finally, instead of focusing on the effect of advertising on a single brand, firm or product category, this paper mainly discusses the aggregate effects of advertising on price indexes and the distribution of firm market shares. Taking the proliferation of internet advertising as an exogenous shock, this paper explores the macroeconomics implications of this shock on the dynamics of aggregate markup and concentration. In summary, this paper analyzes a novel and comprehensive data set with a structural approach, in hopes of understanding the *macroeconomic* effects of advertising.

To illustrate the main mechanisms of the quantitative model, I first study a one-sector economy with a single representative household and N heterogeneous firms. Firms each produce a differentiated product with heterogeneous production costs, and compete in a monopolistic competitive market. Each firm can either cut prices or invest in advertisements to attract higher demands. The representative household has brand preferences towards each firm's products, and its brand preferences are endogenously determined by the household's exposure to each brand's advertisements.

There are two main results I derive from this one-sector model. First, advertising creates two general equilibrium effects on demand, which I label as "quality" and "price" effects. The quality effect is the direct demand response through higher brand preferences (i.e. higher perceived qualities). The price effect is the indirect impacts of a firm's advertisements on the price indexes of the product category it belongs. While the quality effect is strictly positive, the price effect can be either positive or negative depending on the advertiser's market share. In a special case when all firms are identical, the economy has a symmetric equilibrium where all firms have equal market shares and the price effect of advertising is zero. This result no longer holds true when firms have heterogeneous production costs.

The second result is that exogenous shocks to the advertising costs can create complicated dynamics to the distribution of firm market shares, depending on the number of firms and their respective production technologies. In some cases, shocks to advertising costs can generate movements of markup and concentration in opposite directions. Similar results are documented by a number of empirical studies, as Syverson (2019) points out, where reductions in transportation or search costs prompt customers to shift towards larger, lower-cost sellers, creating higher concentration but lower markup. In our model, reductions in advertising costs can raise markup and lower concentration, because cheaper ads allow smaller, less-efficient firms to secure greater shares of the market. The underlying logic of our result is consistent with previous studies on related topics.

Note that a key assumption of our model is that firms compete for brand preferences in a "zero-sum" way: holding prices constant, if all firms increase their advertising spending such that the household's proportional exposures to each brand stay the same, then the household's consumption does not change. Previous literature finds supporting evidence for this assumption. For example, Hartmann and Klapper (2015) show that Super Bowl ads can generate significant increase in demand for soda brands, but much of this gain diminishes if two major soda brands both advertise. The zero-sum assumption also justifies the minuscule demand effect of advertising found in many empirical studies. Spending more on advertisements does not promise a surge in product demand, especially when other competitors also advertise more aggresively. After all, advertisements aim to compete for consumer's time and purchasing power, two resources with finite and sometimes inelastic supply. While advertisers can double or triple their marketing budgets, consumers can rarely increase viewership or consumption by the same factor, due to their time and budget constraints.

The second part of this paper presents empirical evidence on the change of market structure and advertising costs from 2010 to 2016, using the firm-level panel data I create. This paper first documents changes to the market structrure during the sample period. Our sample covers 453 narrowly defined product categories, known as product modules, including goods commonly sold in grocery and drug stores such as food and beverages, cosmetic, toys, and et cetera. I then measure Herfindahl index and markup at aggregate as well as product module level, using demand-side estimation approach following Hottman, Redding and Weinstein (2016). I find that between 2010 and 2016, aggregate Herfindahl decreases about 10% while aggregate markup decreases for around 5%. At product module level, only 46% of the 453 product modules have the same aggregate trends of decreasing markups and Herfindahls, while 30% of product modules have markups and Herfindahls changing in opposite directions. These findings suggest that aggregate statistics may not provide complete pictures for the change in U.S. market structures over time.

This paper then proceed to document changes in the aggregate cost function of advertising. We use the average and marginal cost per impression as measures for the expensiveness of advertisements. During the sample period, the average costs of advertising have been increasing, while marginal costs have been decreasing. The "impression elasticity" of advertising, which measures the percentage increase in viewership in response to a 1% increase in advertising spending, rises from 0.94 in 2010 to 0.99 in 2016. The aggregate cost function for the advertising industry is becoming closer to constant returns to scale, which is either driven by the proliferation of online advertising or the alterations in the pricing models of other advertising platforms. I also find that the changes in advertising marginal costs vary across advertisers with different sizes, using quantile regression methods a la Koenker and

Hallock (2001). I show that the reduction in marginal cost is the greater for advertisers above median. The share of advertisers with increasing marginal returns from advertising also increased from 2010 to 2016.

In the final section of this paper, I formally develop a quantitative model with multi-product, multi-sector heterogeneous firms. The main framework is similar to Hottman, Redding and Weinstein (2016), but we allow firms to endogenously choose advertisement levels. I first show that for firms with positive advertising spending at the equilibrium, each firm's marginal revenue from advertising is equal to its demand elasticity. This result is a generalization of the classic finding in Dorfman and Steiner (1954), but in a framework with multi-product, heterogeneous firms. Another therotical contribution of this paper is to derive the partial-equilibrium relationship between advertising expenditure and product entry, both being endogenous variables of this model. I show that a firm's profit gain from introducing a new product is positively correlated with its brand preferences. If a firm arbitrarily increases its advertising spending to attain higher level of brand preferences, then it becomes more profitable for the firm to introduce new products. I then structurally estimate the model parameters using firm-level data on advertising spending, prices and market shares, and show that the effectiveness of advertising on brand preferences is greater in product categories with lower elasticities of substitution. Finally, I assume the cost structure of advertising in 2016 stayed the same as in 2010, and construct counterfactual distributions of firm market shares under this hypothetical scenario. The counterfactual analysis shows that the aggregate markup and concentration would both rise from 2010 to 2016, had the advertising technology stayed the same during this period.

Literature. This paper connects multiple strands of literature in macroeconomics, industrial organization and marketing. First, it is related to a growing number of studies that document the evolution of market power in the US. For example, Neiman and Vavra (2019) documents that aggregate concentration has declined by 20% from 2004 to 2015, despite that household-level concentration has increased. Different from their study, this paper uses a more comprehensive scanner dataset and measures concentration at the firm level instead of product level. Our results suggests that aggregate concentration has decreased by 10%

between 2010 to 2016, which supports the findings in Neiman and Vavra (2019). On the aggregate trend of markup, De Loecker, Eeckhout and Unger (2020) show that the average markup in the US has been steadily rising between 1980 and 2016, from 21% to 61%. In this paper, I find that aggregate markup has been decreasing by 5% between 2010 to 2016, which seems at odds with a growing body of empirical evidence on this subject (Hall 2018, Traina 2018).

Why do our results differ? The reason could be differences in estimation methods and data sources. Most macroeconomic studies on this topic adopt the "supply-side" approach to estimate markups, following Hall (1984), De Loecker and Warzynski (2012) and De Loecker et al. (2016). This paper, on the other hand, uses the so-called "demand-side" approach, following Berry, Levinsohn and Pakes (1995), Goldberg (1995), Feenstra and Weinstein (2010), and more recently Hottman, Redding and Weinstein (2016). Different from the supply-side approach, the demand-side approach makes explicit assumptions about consumer preferences and the competitive environment, which are necessary in our case to study the effect of advertising on market structures. I also use a different dataset from De Loecker, Eeckhout and Unger (2020), and mainly focus on consumer packaged goods sold in grocery stores and supermarkets. While this dataset covers fewer industries and exclude consumption in certain important categories such as automobiles, education and housing, our sample includes more small firms than Compustat, which only includes publicly traded firms. To summarize, this paper is the first to use demand-side estimation approach to document changes in aggregate markup, to the best of my knowledge.

This paper is also related to a rapidly growing literature that studies the role of customer markets in a macroeconomic context, dating back to Phelps and Winter (1970) and Klemperer (1995). For example, Ravn et al. (2006) and Nakamura and Steinsson (2011) study the effect of product-level consumption habit formation on firm's price-setting behaviors. Gourio and Rudanko (2014) develop a search theoretic model with frictional matching between consumer and firms. Recent papers in this literature also feature heterogeneous firms, where the heterogeneity originates from financial shocks (Gilchrist et al. 2017) or productivity shocks (Paciello, Pozzi and Trachter 2018). Our result of decreasing aggregate markup provides some supporting evidence for the customer market models, which usually imply counter-cyclical markups. But more importantly, this paper addresses the critiques raised by Hall (2014) and Fitzgerald and Priolo (2018), where the authors point out that fluctuations in markup alone can not justify the pro-cyclicality of advertising spending or the changes in firm market shares. Our paper resolves this issue by proposing a model where firms can either cut prices or spend in advertising in order to compete for higher market shares.

A vast literature in industrial organization and marketing focuses on the economic effects of advertising, as surveyed by Bagwell (2007). Different from many studies in this literature, my paper does not address the theoretical debate on whether the role of advertising is informative, persuasive or complementary. In our model, firms use advertising to compete for higher demand, conditional on their prices. The additional demand can result from differences in actual quality, "perceived" quality, or simply taste. This paper does not take a stand on the exact mechanism through which advertisements generate consumer brand preferences – this question is beyond the scope of the current project*. This paper is also related to a number of studies that use supermarket scanner datasets to explore the effect of advertising (Ackerberg 2003) or the persistence of brand preferences (Bronnerberg, Dube, and Gentzkow 2012). Finally, Dinlersoz and Yorukoglu (2012) study the theoretical implication of declining cost of information dissemination to firm and industry dynamics; Molinari and Turino (2017) explore the effect of aggregate advertising on aggregate consumption using a DSGE model. This paper studies a similar research question, even though it uses different data sources and estimation approaches.

Layout. The rest of this paper is structured as follows. Section 2 presents a simplified one-sector model to illustrate the main mechanisms of our full model. Section 3 describes the data sources as well as the steps to combine data. Section 4 provides empirical evidence on markup, concentration and advertising cost structures. Section 5 presents the full quantitative model, estimates model parameters from data, and discusses results from counterfactual analysis. Section 6 concludes.

2 Optimal Advertising Strategy in a One-Sector Model

I start by introducing a one-sector model to illustrate the effect of advertising on product demand and market structure. In this stylized model, the economy consists of N homogeneous firms and a representative household. The firms choose prices under Bertrand competition, but in addition can use advertising to influence the household's "brand preferences" – propensity to purchase certain brands over others, even when prices for the desired brands are identical or higher than the alternatives.

The stylized model serves two purposes. First, it illustrates the partial and general equilibrium effects of advertising in a static, one-sector economy. Second, the model shows how innovations in advertising technology alters the distribution of market shares – measured by aggregate concentration and markups – when firms have heterogeneous production costs. This simplified framework illustrates the main mechanisms and results from our quantitative model, while abstracting away from additional details concerning multi-brand firms and product hierarchies.

There are several results from the one-sector model. First, I show that in equilibrium the marginal revenue gain of advertising equals to the elasticity of demand, for any firms that spend positive amounts on advertising. This is a classical result that dates back to Dorfman and Steiner (1954). Second, when firms have identical production technologies, there exists a symmetric equilibrium where all firms charge the same price, advertise for the same amount, and each secures an equal share of the market. Third, the general equilibrium effect of advertising on demand can be decomposed into a "quality" effect (changing brand preferences) and a "price" effect (changing the aggregate price index). In a symmetric equilibrium with identical firms, the price effect of advertising is exactly zero. When firms are heterogeneous, however, changes in advertisement levels can generate changes in the aggregate price index, and further create welfare impacts to the representative household. Finally, technological innovations that alter the cost structure of advertising can reshape the distribution of market shares, when firms are heterogeneous. Most surprisingly, such redistribution of market shares can cause markups and concentration to move in *opposite* directions - that is, the market becomes less concentrated but markup gets higher, or vice versa - as a result from improved advertising technology over time.

2.1 Demand

A representative household of unit measure has the following preferences over products from N differentiated brands:

$$u(c_1, c_2, \ldots, c_N) = \left[\sum_{i=1}^{N} \left(\frac{\varphi_i}{\tilde{\varphi}} c_i\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

The term φ_i represents the household's "brand preferences" in brand i – that is, the additional utility obtained from consuming products from brand i, as a result of either quality, taste differences or influences from advertising. Note that the brand preferences φ_i are divided by the geometric mean $\tilde{\varphi} \equiv \left(\prod_{i=1}^{N} \varphi_i\right)^{1/N}$, so that the utility function only accounts for the household's *relative* tastes for each brand, not the absolute levels. In other words, if all φ_i are multiplied by the same constant, holding prices the same, the household's total utility will not change.

For simplicity, I normalize both the geometric mean of brand preferences $\tilde{\varphi}$ and the household's total income to 1. The household's problem then becomes:

$$U = \max_{c_i} \; \left[\sum_{i=1}^{N} (\varphi_i c_i)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$
$$\text{s.t.} \; \sum_{i=1}^{N} p_i c_i = 1$$

2.2 Production

There are N firms in the economy, and each firm owns a differentiated brand[2]. Each firm i can supply its products through a linear production function, with constant marginal cost θ_i. Unlike the standard model of Bertrand competition with differentiated goods, here firms compete with both prices and advertisements. More specifically, the household's brand preferences φ_i are determined by the following equation:

$$\frac{\varphi_i}{\tilde{\varphi}} = \frac{q(\eta_i)}{\left(\prod_{i=1}^{N} q(\eta_i)\right)^{1/N}} \tag{2}$$

where η_i is the advertising expenditure of brand i, and q is defined as the advertising *impression* function, which is a mapping between the dollar amounts spent on ads and the number of viewers (impressions) the ads reach. We assume that q is positive, strictly increasing and concave on $[0, \infty)$. In addition, we impose the assumption that $q(0) > 0$. This assumption guarantees that even a brand does not advertise at all, its impression does not drop to zero[3]. In sum, firm i's profit is given by the following expression, where prices $\boldsymbol{p_{-i}}$ and advertising levels $\boldsymbol{\eta_{-i}}$ of its competitors are taken as given:

$$\pi_i(p_i, \eta_i | \boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}}) = (p_i - \theta_i)c_i(p_i, \boldsymbol{p_{-i}}, \eta_i, \boldsymbol{\eta_{-i}}) - \eta_i \tag{3}$$

here $c_i(p_i, p_{-i}, \eta_i, \eta_{-i})$ is the household's demand on brand i.

2.3 Equilibrium

Definition 1 *(Equilibrium)*

An equilibrium is defined as a set of consumption levels $\boldsymbol{c}^* \equiv \{c_1^*, c_2^*, \ldots, c_N^*\}$*, prices* $\boldsymbol{p}^* \equiv$

[2]In this simplified model, the notions of "firms" and "brands" are interchangeable, because each firm owns only one brand. In the full model, a brand is defined by the intersection of a firm and a product category.

[3]This guarantees that the marginal utility of each brand's product is always positive, regardless of its advertising levels.

$\{p_1^*, p_2^*, \ldots, p_N^*\}$ *and advertising levels* $\boldsymbol{\eta}^* \equiv \{\eta_1^*, \eta_2^*, \ldots, \eta_N^*\}$ *such that:*

(i) Given prices $\boldsymbol{p}^*$ *and advertising levels* $\boldsymbol{\eta}^*$*, the representative household chooses consumption bundle* $\boldsymbol{c}^*$ *to maximizes its utility, subject to the budget constraint;*

(ii) Each firm i *maximize its profit by choosing prices* $\tilde{p}_i(\boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}})$ *and advertising level* $\tilde{\eta}_i(\boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}})$ *as a best response to its competitor's strategy* $\{\boldsymbol{p}_{-i}, \boldsymbol{\eta}_{-i}\}$*;*

(iii) Each firm's strategy is the best response to the other firm's strategy: $p_i^* = \tilde{p}_i(\boldsymbol{p^*_{-i}}, \boldsymbol{\eta^*_{-i}})$ *and* $\eta_i^* = \tilde{\eta}_i(\boldsymbol{p^*_{-i}}, \boldsymbol{\eta^*_{-i}})$*, for* $i = 1, 2, \ldots, N$*.*

To solve for the equilibrium, I first derive the demand functions from the household's problem. Proposition 1 summarize the demand functions and define the aggregate price index in this framework. This demand system is closest to the single-nested CES demand in Redding and Weinstein (2019), except that in my model the demand shifters φ_i are defined as "brand preferences" and are determined endogenously determined by firm's advertising expenditures.

Proposition 1 *The household's demand for product* i*, as a function of product prices* $\boldsymbol{p}$ *and advertising expenditures* $\boldsymbol{\eta}$*, is given by:*

$$c_i(\boldsymbol{p}, \boldsymbol{\eta}) = \frac{p_i^{-\sigma} q(\eta_i)^{\sigma-1}}{\sum\limits_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}} = \frac{p_i^{-\sigma} \varphi_i^{\sigma-1}}{\sum\limits_{j=1}^{N} p_j^{1-\sigma} \varphi_j^{\sigma-1}} = \frac{(p_i/\varphi_i)^{1-\sigma}}{P^{1-\sigma}} \frac{1}{p_i} \tag{4}$$

where the aggregate price index is defined as

$$P = \left[\sum_{i=1}^{N} \left(\frac{p_i}{\varphi_i} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{5}$$

One special example is when $\varphi_i = 1$ for all $i = 1, \ldots, N$, in which case the demand system is identical to a standard constant elasticity of substitution (CES) model. When the household does not prefer any particular brand over others, the market share of each product is a function of the ratio between the products's own price and the price index. However, if the household develops brand preferences toward some brand k (i.e. $\varphi_k > 1$), the result on brand

k's market share is equivalent to a reduction of its own price p_k and a proportional increase in its competitors' prices p_j, for all $j \neq k$. Corollary 1 provides the expression of each brand i's market shares, in the general case when $\varphi_i \neq 1$.

Corollary 1 *The household's expenditure share on brand i is:*

$$S_i = \frac{p_i^{1-\sigma} q(\eta_i)^{\sigma-1}}{\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}} = \frac{p_i^{1-\sigma} \varphi_i^{\sigma-1}}{\sum\limits_{j=1}^{N} p_j^{1-\sigma} \varphi_j^{\sigma-1}} = \frac{(p_i/\varphi_i)^{1-\sigma}}{P^{1-\sigma}} \tag{6}$$

When firms have identical marginal cost of production, Proposition 2 shows the closed-form solutions of the equilibrium outcomes. The full proof is included in the mathematical appendix, where I solve the best response functions for each firm, and show that a unique symmetric equilibrium exists for this game. Depending on the functional form of $q(\cdot)$, each firm's optimal advertising expenditure can be either zero or positive in the symmetric equilibrium.

Proposition 2 *If all firms have identical production technology, i.e. $\theta_i = \theta_j = \theta$ for any $i \neq j$, then there exists a unique symmetric equilibrium where $p_i = p^*$ and $\eta_i = \eta^*$ for all $i = 1, 2, \ldots, N$, such that:*

$$p^* = \frac{1 + (N-1)\sigma}{(N-1)(\sigma-1)} \theta$$

$$\eta^* = \begin{cases} 0 & \textit{if } \frac{q'(0)}{q(0)} < \frac{1+(N-1)\sigma}{(N-1)(\sigma-1)} N \\ f^{-1}\left(\frac{1+(N-1)\sigma}{(N-1)(\sigma-1)} N\right) & \textit{if } \frac{q'(0)}{q(0)} \geq \frac{1+(N-1)\sigma}{(N-1)(\sigma-1)} N \end{cases}$$

where $f(\eta) = \frac{q'(\eta)}{q(\eta)}$ and $f^{-1}(\cdot)$ is its inverse function.

When firms are heterogeneous, it is difficult to derive the closed-form solutions for the equilibrium. Still, it is possible to analyze the relationship between optimal advertising spendings and other variables such as prices and market shares. Proposition 3 states a general result on the relationship between optimal advertising levels and optimal prices.

Proposition 3 *(Dorfman and Steiner) In an equilibrium with positive advertising spending* *for some firm i, the marginal increase in firm i's revenue from advertising is equal to its elasticity of demand:*

$$p_i \frac{\partial c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial \eta_i} = \epsilon^D_{i,p} \equiv -\frac{\partial c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial p_i} \frac{p_i}{c_i(\boldsymbol{p}, \boldsymbol{\eta})}$$

Intuitively, a firm i has two ways to promote its products, either by 1) spending additional ϵ dollars on advertisements or 2) cutting prices by ϵ/c_i. The two means of promotion are equally costly (when ϵ is small), so in the equilibrium the marginal benefit from both methods must be equal. This classical result dates back all the way to Dorfman and Steiner (1954), where the authors come up with the original argument without assuming any specific forms for the demand function. Becker and Stigler (1977) also provide similar intuitions in their seminal paper. I show in the quantitative part of my paper that a more general version of this result also holds true when each firm owns multiple brands and multiple products.

2.4 Discussion

In this section, I discuss in greater details about the results and implications from the stylized model. I focus on the key mechanisms through which advertising affects demand, markup and prices, and show that changes in the cost structure of advertising have real impacts on the economy. Numerical results show that in the heterogeneous firm case, improvements in the advertising efficiency reshape the equilibrium firm size distribution, creating changes in aggregate markup and market concentration.

2.4.1 Markup and Market Shares

Before I start analyzing the general equilibrium effect of advertising, it is important to first understand the firms' optimal pricing rule. Similar to results from Hottman, Redding and Weinstein (2016) and more recently Neiman and Vavra (2019), the equilibrium in this framework features variable firm-level markups that depend on the firm's market share. To see this, take logarithm of the demand function from (4):

$$\log c_i(\boldsymbol{p}, \boldsymbol{\eta}) = (-\sigma) \log p_i + (\sigma - 1) \left[\log \varphi_i(\eta_i) + \log P\right] \tag{7}$$

where P is the aggregate price index. The demand elasticity of product i is therefore:

$$\Rightarrow \quad \epsilon_{i,p}^{D} \equiv -\frac{\partial \log c_i}{\partial p_i} p_i = \sigma - (\sigma - 1)\frac{\partial \log P}{\partial p_i} p_i \tag{8}$$

The last term in Equation 8 represents the "externality" that each firm's pricing decision imposes on the aggregate price index. In a traditional Dixit-Stiglitz demand system, the number of firms in the market is assumed to be large, so that this externality on aggregate price index is negligible. In that case, the elasticity of demand $\epsilon_{i,p}^{D}$ is equal to the (constant) elasticity of substitution σ. In the current model, I drop the assumption that the number of firms is large, thus allowing each firm to internalize the consequences of its own pricing decisions on the aggregate price index. It turns out that the magnitude of this second-order effect is proportional to the market share of firm i:

$$\frac{\partial \log P}{\partial p_i} p_i = \frac{1}{1-\sigma}\left[\sum_{j=1}^{N}\left(\frac{p_j}{\varphi_j}\right)^{1-\sigma}\right]^{\frac{\sigma}{1-\sigma}} (1-\sigma)\left(\frac{p_i}{\varphi_i}\right)^{1-\sigma}\frac{1}{P} \tag{9}$$

$$= \frac{\left(\frac{p_i}{\varphi_i}\right)^{1-\sigma}}{\sum_{j=1}^{N}\left(\frac{p_j}{\varphi_j}\right)^{1-\sigma}} = S_i \tag{10}$$

Combining the results, Proposition 4 summarizes the optimal pricing rule. Note that instead of a constant markup over marginal cost, each firm's markup increases with its market share.

Proposition 4 *(Markup) The equilibrium pricing rule for firm i is characterized by the following equation:*

$$\mu_i \equiv \frac{p_i}{\theta_i} = \frac{\epsilon_{i,p}^{D}}{\epsilon_{i,p}^{D} - 1} \tag{11}$$

where the demand elasticity for firm i is given by:

$$\epsilon_{i,p}^{D} = \sigma - (\sigma - 1)S_i \tag{12}$$

I want to make two remarks here. First, there is no markup dispersion unless firms have heterogeneous production costs. If all firms are identical, then at the equilibrium each firm secures an equal share of the market, and the elasticities of demand for all firms are:

$$\epsilon_{i,p}^{D*} = \frac{1 + (N-1)\sigma}{N} = \sigma - \frac{\sigma - 1}{N}$$

The markups for all firms are therefore identical:

$$\mu^* = \frac{\epsilon_{i,p}^{D*}}{\epsilon_{i,p}^{D*} - 1} = \frac{1 + (N-1)\sigma}{(N-1)(\sigma-1)} = \frac{\sigma}{\sigma - 1} + \frac{1}{(N-1)(\sigma-1)} \tag{13}$$

Second, when firms are homogeneous and the number of firms is fixed (no entry and exit), advertising has no impact on aggregate markup (and price levels) at all.

2.4.2 General Equilibrium Effects of Advertising

Let's return to the log-transformed demand function in equation (7), and analyze the general equilibrium effect of advertising spending on demand. The first order derivative with respect to advertising spending η_i is:

$$\frac{\partial \log c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial \eta_i} = (\sigma - 1)\frac{\partial \varphi_i(\eta_i)}{\partial \eta_i} + (\sigma - 1)\frac{\partial \log P}{\partial \eta_i} \tag{14}$$

Write in elasticity terms:

$$\epsilon_{i,\eta}^{D} = \underbrace{(\sigma-1)\epsilon_{i,\eta}^{\varphi_i}}_{\text{quality effect}} + \underbrace{(\sigma-1)\epsilon_{i,\eta}^{P}}_{\text{price effect}} \tag{15}$$

where $\epsilon_{i,\eta}^{\varphi_i} \equiv \frac{\partial \varphi_i}{\partial \eta_i}\frac{\eta_i}{\varphi_i}$ and $\epsilon_{i,\eta}^{P} \equiv \frac{\partial P}{\partial \eta_i}\frac{\eta_i}{P}$ are the elasticities of (1) brand preferences (2) aggregate price index with respect to advertising spending of brand i. The assumption $q'(\eta) > 0$ and the definition of brand preferences in equation 2 guarantees that the first term in equation 15 is always positive. Intuitively, when firms increase their advertising expenditures, holding all other variables constant, they raise their brand preferences (perceived quality by household), causing demand to increase. I name this term the "quality effect" of advertising. In addition, advertising also triggers a second-order, general equilibrium effect on demand through the aggregate price indexes, which can be either positive or negative. This is the "price effect" of advertising.

To further decompose the price effect, plug in the definition of aggregate price index from equation 5 and rewrite $\epsilon^{P}_{i,\eta}$ as:

$$\begin{aligned}
\epsilon^{P}_{i,\eta} &= \frac{\partial}{\partial \eta_i}\left[\frac{1}{1-\sigma}\log\left(\sum_{j=1}^{N}\left(\frac{p_j}{\varphi_j}\right)^{1-\sigma}\right)\right]\eta_i \qquad (16)\\
&= -\left(\frac{\left(\frac{p_i}{\varphi_i}\right)^{1-\sigma}}{\sum_{j=1}^{N}\left(\frac{p_j}{\varphi_j}\right)^{1-\sigma}}\right)\left(\frac{\partial \varphi_i}{\partial \eta_i}\frac{\eta_i}{\varphi_i}\right) - \sum_{k\neq i}\left(\frac{\left(\frac{p_k}{\varphi_k}\right)^{1-\sigma}}{\sum_{j=1}^{N}\left(\frac{p_j}{\varphi_j}\right)^{1-\sigma}}\right)\left(\frac{\partial \varphi_k}{\partial \eta_i}\frac{\eta_i}{\varphi_k}\right)\\
&= \underbrace{-S_i\epsilon^{\varphi_i}_{i,\eta}}_{\text{spillover effect}} - \underbrace{\sum_{k\neq i} S_k\epsilon^{\varphi_k}_{i,\eta}}_{\text{competition effect}}
\end{aligned}$$

When a firm raises its advertising budget, it creates two types of externalities on its competitors, through changes in the aggregate price index. The first is a "spillover" effect – that is, a positive externality on the demands of all rival goods sold in the same category. Previous studies find empirical evidence of such positive spillover effects in advertisements of antidepressants (Shapiro (2018)), restaurants (Sahni (2016)) and a number of other categories (Lewis and Nguyen (2012)). The second is a "competition" effect, which is a negative externality to rival demands. When firm i increases advertising spending, household's brand preferences in all other firms declines proportionately, holding prices constant. In other words, because the utility function is homogeneous of degree 0 in brand preferences, the competition of advertising is zero-sum in nature: one firm's gain is all other firms' loss.

As suggested by equation (16), the magnitude of spillover and competition effects depend on the market shares (S_i) and impression elasticities (own elasticity $\epsilon^{\varphi_i}_{i,\eta}$ and cross elasticities $\{\epsilon^{\varphi_j}_{i,\eta}\}_{j\neq i}$). In a symmetric equilibrium, the spillover and competition effects exactly offset each other, which means that the price effect is zero, and the net advertising elasticity $\epsilon^{D}_{i,\eta}$ should always be positive. When there exists firm heterogeneity in the production technology, firms have different market shares and therefore the price index effect generally does not equal to zero. In fact, if a firm's market share is greater than the average market shares of its competitors, the spillover effect of its advertisements would exceed the competition effect, causing the overall price effect to be negative. In the extreme case when a firm's market share is very close to 1, the price effect can be so large that the total advertising elasticity of demand becomes negative. For this firm, advertising does more harm than good, so the optimal level of advertising would be no advertising at all.

2.4.3 Comparative Statics: Innovation in Advertising Technologies

So far the discussion mainly focuses on the equilibrium allocations for prices and advertising levels, with the advertising technology $q(\cdot)$ taken as given. In this section, I explore the effect on equilibrium outcomes when the advertising techonology changes exogenously. Over the last two decades, the proliferation of online and mobile advertising has foundamentally changed the marketing industry. Add some statistics here. Probably the most pronounced change is in the cost structure of advertising. In traditional TV advertising, for example, the technology is usually decreasing returns to scale: TV commercials that are ten times more expensive does not reach ten times more viewers. Online advertising, however, commonly applies a pay-per-view or pay-per-click pricing model, which is inherently constant returns to scale in terms of impressions. To capture these different advertising technologies, I assume the following functional form for the advertising impression function:

$$q(\eta_i) = \left(\lambda + \frac{\eta_i}{k}\right)^{\beta} \tag{17}$$

Roughly speaking, k and β captures the average and marginal expensiveness of advertising. The parameter $\lambda > 0$ guarantees that the assumption $q(0) > 0$ holds for all values of η_i. Note that when λ is small, β is also an approximate measure for the returns to scale of advertising technology. The question of interest is how changes in k and β affect the distribution of firm market shares in the equilibrium. Of course, the first step is to find an algorithm to numerically calculate the equilibrium outcomes, when firms have heterogeneous marginal costs. I use the following iterative algorithm to find the equilibrium:

Algorithm 1 (Nash Equilibrium with Heterogeneous Firms)
The following algorithm finds the equilibrium allocation $X^* = (\boldsymbol{p^*}, \boldsymbol{\eta^*})$*, when firms have heterogeneous marginal costs:*

1. *Label firms with random indexes* $1, 2, \ldots, N$.
2. *Start with an initial guess* $X_0 = (\boldsymbol{p_0}, \boldsymbol{\eta_0})$ *for all firms' optimal strategies.*
3. *Starting from firm 1, find its best response* $(p_{1,1}, \eta_{1,1})$ *given the strategy of its opponents* $(\boldsymbol{p_{-1,0}}, \boldsymbol{\eta_{-1,0}})$ *by maximizing its profit function in equation* (3)*. Update* $X_1 = (p_{1,1}, \boldsymbol{p_{-1,0}}, \eta_{1,1}, \boldsymbol{\eta_{-1,0}})$.
4. *Repeat Step 3 for firm 2, using* X_1 *as the starting point. Update* X_2 *similarly.*
5. *Repeat Step 3 recursively for all firms in the same order. Stop at iteration* T *if for any* $j = 1, 2, \ldots, N$*:*
$$|X_{T-j} - X_{T-j-1}|^2 < \epsilon$$
where ϵ *is the chosen tolerance level.*

Figure 1 and 2 show the computational solutions for two sample economies. In both cases, I start with a fixed number of firms, and the firms' marginal costs are drawn randomly from a normal distribution $N(\mu_\theta, \sigma_\theta)$. I then use Algorithm 1 to find the Nash equilibrium under different values of k and β. Depending on the distribution of firm marginal costs $\boldsymbol{\theta}$, the model can generate many types of market structures in the equilibrium. Figure 1 and 2 provide two examples to illustrate this point. The top panels in Figure 1 and 2 feature equilibria with duopoly firms; the bottom panels show equilibria with more than two firms, and when firms endogenously choose to exit the market under some parameter values of k and β. We see that in the duopoly case, improvements in the advertising technology changes the aggregage Herfindahl and markup in opposite directions. When the equilibria consist of more than two firms, the comparative statics results can become more complicated. Even though total advertising spending always increase when advertisements get cheaper(lower k or higher β), the implications on equilibrium firm share distribution is less clear. On one hand, lower **average** costs of advertising (lower k) promote market entry, reduce concentration but raise markup. On the other hand, lower **marginal** costs of advertising (higher β) deters entry, raise both concentration and markup.

In summary, depending on the heterogeneous production costs, this model can generate either positively or negatively correlated movements in Herfindahl index and markup, as costs of advertising change over time.

3 Data Description

In this section, I briefly describe the data sources and the method to combine them into a firm-level panel data set.

3.1 Advertising Expenditure: Nielsen Ad Intel Data

Nielsen Ad Intel provides occurrence-level advertising information such as time, duration, format, and expenses paid for each advertisement. The data is available for ads featured on TV, internet, radio, newspaper, magazines and other media platforms from 2010 to 2016. As Table 1 shows, the total advertising revenue included in the Ad Intel from 2010 to 2016 is approximately $108 billion. Around 60% of total advertising revenue comes from network TV ads, 26% from magazine ads, 6% from spot (local) TV ads, and the remaining 7-8% from other types of advertisements such as radio, outdoor, internet, newspaper, and free standing insert (FSI) ads.

For this project, I select advertisements for goods sold in grocery and drug stores, including food and beverages, cosmetics, tobacco, toys, pet supplies, soaps and other cleaners. I exclude advertisements for goods not commonly sold in grocery stores, such as automobiles, since prices and quantities of these goods are not observed. I also ignore advertisements in "medicines and remedies" category, as the demands of goods featured in these ads are usually determined by factors such as personal health conditions and doctor approvals, not prices or brand preferences.

3.2 Grocery Sales: Nielsen Scanner Datasets

For grocery sales information, I use two scanner datasets (Nielsen Household Panel and Nielsen Retailer Scanner). Here "scanner" datasets refer to the data-collection method: information in these datasets is collected from barcode scanners, either at grocery stores with point-of-sales (PoS) systems, or at home with handhold devices.

3.2.1 Nielsen Retailer Scanner Data

Nielsen Retail Scanner Data, also known as Kilts-Nielsen Retailer Scanner (KNRS) or RMS, is a store-level panel dataset containing weekly sales information for over 35000 stores across the U.S. from 2006 to 2016. The data reports weekly price and quantity information for each product with a UPC (Universal Product Code) barcode sold at each covered store, collected from in-store PoS systems.

One main advantage of KNRS is its broad coverage. Table 2 provides some summary statistics of the dataset. As shown in the table, KNRS contains more than 13 billion transaction records worth more than $220 billion each year. According to previous studies, this dataset represents around 30 percent of total U.S. expenditure on food and beverages and 53 percent of all sales in grocery stores[4].

3.2.2 Nielsen Household Panel Data

Nielsen Household Panel Data, also known as Kilts-Nielsen Consumer Panel (KNCP), Nielsen Homescan or simply HMS data, is a household-level longitudinal panel from 2004 to 2016. Each year, around 60,000 panelists across the U.S. provide detailed information on their grocery trips, using handhold barcode scanning devices provided by Nielsen. After each grocery trip, households enter date and store information of the trip on their devices, and scan barcodes for all goods they purchase. If the trip is made to a Nielsen-partnered store, price of

[4]For example, see Beraja et. al (2016) and Argente et al.(2018).

each item is automatically set to the average price at that store during the week of purchase. In other cases, panelists are required to record prices from the receipts manually.

To encourage accurate and timely reporting, Nielsen provides incentives such as sweepstakes, prize drawings and gift points to panelists each month. Households must transmit purchases exceeding a minimum dollar amount to stay "active" each month, and are only included in the panel if they stay active for all 12 months in a year. Overall, attrition rate of panelists is around 20% annually. New panelists are recruited each year to replace exiting ones, to keep the size of the panel unchanged.

Table 3 shows some summary statistics of the Household Panel Data.

3.3 Matching UPCs with Firms: GS1 US

An important data source that allows me to match each product barcode with its manufacturer is the GS1 US database. GS1 is the non-profit organization responsible for registration and maintenance of all UPC barcodes worldwide. The GS1 US database provides firm-level administrative information for more than 450,000 companies around the world. Users are able to link grocery products with their manufacturing companies using the first 6-10 digits of UPC barcodes (also known as the "company prefix"). While most firms only have a single company prefix, some larger firms own multiple ones due to previous mergers and acquisition. For more details about this dataset, readers can refer to Hottman, Redding and Weinstein (2016) as well as Argente, Lee and Moreira (2018).

3.4 Cleaning Data

In this section, I describe the procedures to clean and merge the four datasets.

First, I link all product barcodes from KNRS to their manufacturers, using company prefixes from GS1. Next, I link firm names in GS1 with the list of advertisers in Nielsen Ad Intel using an approximate string matching (also called fuzzy word matching) algorithm. Finally, we calculate monthly sales revenue and advertising spendings to create a monthly panel data at firm level. Table 4 shows some summary statistics of the cleaned data.

The second step above needs some further discussion. When matching firm-level data across Nielsen Ad Intel and KNRS, we notice that many firm names appear differently across two databases, such as “Pepsi-Cola North America Inc.” and “ PEPSICO INC”. Pairing these names manually is difficult, as there are 65,000 × 10,000 potential pairs of firm names to match. To solve this problem, we find and remove common company suffixes (such as ”Inc”, ”Co”, ”Ltd”, etc.) from firm names[5]. Next, we measure the longest common substring (LCS) distances between firm name pairs to determine their similarities. Finally, we fine tune the matching results by choosing a maximum string distance threshold, in order to match same firms with slightly different names (such as “Hershey Co.” and “The Hershey Company”) instead of different firms with similar names (such as “3M” and “IBM”).

4 Empirical Analysis

Next, I briefly overview the main empirical results found in our data set.

I first document changes of the market structure from 2010 to 2016, measured by Herfindahl index and markups. The empirical evidence suggests that over theis period, both aggregate concentration and markup have been declining. At the product-category level, our analysis reveals large heterogeneity in the directions of these trends. More than a third of product categories have concentration and markup moving in opposite directions.

From the Ad Intel databse, I find that the average cost to advertise increases while the *marginal* cost decreases over the sample period. A comparison between TV and online advertising shows that the latter is becoming much more expensive on average. Finally, the decrease in marginal costs are largest for medium to large advertisers.

[5]To be exact, we count the number of times each word appears in firm names, and remove 13 words that repeats the most. They are: Inc, LLC, Co, Ltd, Corp, Products, Group, Intl, “The”, Company, Corporation, International, and Enterprises.

4.1 Changes in Market Structure

During the last two decades, the distribution of firm market shares in the U.S. has changed quite dramatically. For example, Neiman and Vavra (2018) document from scanner data that "aggregate Herfindahl has actually *declined* by an average of roughly 20 percent from 2004-2015." Using firm-level Census data, De Loecker, Eeckhout, and Unger (2018) find that average markups increased from 21% in 1980 to 61% in 2016. These findings both suggest that the underlying distribution of firm market shares has changed significantly over time.

However, aggregate statistics can not fully capture the transformation in firm share distribution, when the changes are heterogeneous across industries.

4.1.1 Changes in Herfindahl Index

I define firm-level market shares in product category g and time t as:

$$S_{fgt} = \frac{\sum\limits_{u \in \Omega^U_{fgt}} p_{ut} q_{ut}}{\sum\limits_{f \in F_{gt}} \sum\limits_{u \in \Omega^U_{fgt}} p_{ut} q_{ut}} \tag{18}$$

Where Ω^U_{fgt} represent the set of products manufactured by firm f, and F_{gt} is the set of firms that operate in category g. The Herfindahl index in product category g is defined as:

$$\mathcal{H}_{gt} = \sum_{f \in F_{gt}} (S_{fgt})^2 \tag{19}$$

To compute aggregate Herfindahl, I calculated the weighted average of Herfindahl across all product categories, using revenue as weights:

$$\mathcal{H}_t = \frac{E_{gt} H_{gt}}{\sum_{g \in G} E_{gt}} \tag{20}$$

Figure 3 shows the quarterly series of Herfindahl index from 2010 to 2016. The figure suggests that both average and median Herfindahl index across product categories have been decreasing during the same period.

4.1.2 Changes in Markup

To compute retail markups, I use the demand-side estimation method à la Hottman, Redding and Weinstein (2016), but appled the method to a larger, more comprehensive scanner data set than the household panel data that the original study chooses. The main purpose for using the larger data set is to generate a product-category-level time series of firm markups, for which the household panel data does not suffice.

In this framework, firm markup is derived as:

$$\mu_{fgt} = \frac{\epsilon^D_{fgt}}{\epsilon^D_{fgt} - 1} \tag{21}$$

where the firm's perceived elasticity of demand, ϵ^D_{fgt}, depends on the firm's market share:

$$\epsilon^D_{fgt} = S_{fgt} + \sigma_g(1 - S_{fgt}) \tag{22}$$

The identification of σ_g follows the same argument as in Hottman, Redding, and Weinstein (2016), and is discussed in full length in the quantitative section of this paper. Next, I find the aggregate markup in each category g by taking the cost-weighted average of all firm-level markups, as suggested by Edmond, Midrigan and Xu (2018):

$$\bar{\mu}_{gt} = \sum_{f \in F_{gt}} \frac{\left(\frac{E_{fgt}}{\mu_{fgt}}\right)}{\sum\limits_{k \in F_{gt}} \left(\frac{E_{kgt}}{\mu_{fgt}}\right)} \mu_{fgt} \tag{23}$$

Where E_{fgt} is the revenue of firm f in category g at time t. The cost-weighted average is robust to different ways of aggregation. Therefore, to compute aggregate markup across all product categories, the same formula applies:

$$\bar{\mu}_t = \sum_{g \in G} \frac{\left(\frac{E_{gt}}{\bar{\mu}_{gt}}\right)}{\sum_{g' \in G} \left(\frac{E_{gt}}{\bar{\mu}_{g't}}\right)} \bar{\mu}_{gt} \tag{24}$$

Where E_{gt} is the total revenue of product category g. Figure 4 shows the quarterly time series of aggregate markup as well as the 95% confidence intervals. Note that these confidence intervals are computed from the standard errors of the elasticity of substitution, which I estimate from data. Because markups are nonlinear functions of elasticities of substitution, I use the so-called delta method[6] to construct these confidence intervals. The results suggest that aggregate markup has been steadily decreasing from 2010 to 2016, though the magnitude of this trend is small.

4.1.3 Relationship Between Concentration and Markup

From the analysis above, we see that both aggregate concentration and markup have been decreasing since 2010. Next, I analyze the relationship between changes in markup and Herfindahl at the product category level. In Figure 5, I plot the log differences of markup between 2016 to 2010 against the same log difference of Herfindahl across the same period. Among the 453 product modules included in our sample, only 189 (around 41%) of them display the same aggregate trends of both decreasing markups and Herfindahls. In the remaining 264 product modules, 93 (21%) of them have both increasing markups and Herfindahls, and 171 (38%) have markups and Herfindahls changing in opposite directions. Obviously, aggregate statistics alone do not give us a complete overview of the changes in market structures.

[6] In short, I replace the nonlinear function with its first order Taylor approximation, and use the usual variance formula to compute the confidence intervals.

4.2 Changes in Cost of Advertising

Could changes in the cost of advertising contribute to the trends of declining aggregate markup and concentration? To answer this question, we first need to understand how the cost structure of advertising shifts over time. According to a report from Zenith, a media research company, U.S. adults on average spend 8 hours per day interacting with various types of media, such as TV, radio, magazines, newspaper, and internet. While consumption on all other media types remains relatively stable, internet usage has soared from less than 80 minutes per day in 2011 to over 170 minutes per day in 2019, exceeding the amount of time spent on TV. The transition in media consumption patterns has profoundly changed the cost structures of the advertising industry.

A key difference between TV and internet advertising is their cost functions. Online advertising usually adopts a pay-per-click or pay-per-view pricing model, which intrinsically has constant returns to scale. On the contrary, TV advertising often features decreasing marginal returns, due to competition between advertisers. TV commercials that reach a greater audience, such as Super Bowl or prime times ads, cost more dollars *per view* than advertisements shown in other time slots. This fundamental difference in the cost structures of advertising for the two dominant media types, as well as structural shifts in media consumption behaviors for U.S. adults, alter the aggregate cost function of the entire advertising industry.

In this section, I estimate the average and marginal cost of advertising, using the advertising expenditure and viewership information available in the Nielsen Ad Intel data.

4.2.1 Average Cost of Advertising

Typically, the advertising industry uses the cost per thousand impressions, or CPM (short for "cost-per-mille") to measure the expensiveness of an advertisement. This commonly adopted measure certainly has its advantages. First, it allows us to compare the average cost of advertising across different media types, even though the definition of impressions varies. Second, we can use the time series of average CPM to track the change of advertising expensiveness across time. Figure 6 shows the average CPM for two media types, national TV and internet, from 2010 to 2016. These time series are calculated from the advertising expenditure and impression data pulled from the Nielsen Ad Intel data set.

Two results call for our attention. First, both national TV and internet advertisements have become more expensive since 2010. This result is consistent with previous empirical findings, such as Teixeira (2014). Second, the average CPM for online advertising rose from less than 4 dollars to close to 10 dollars between 2010 and 2012, and stayed at that high level from 2012 to 2016. Additional findings, including Figures X and Y in Appendix Z, suggest that this result is not driven by outliers or data coverage issues. A reasonable hypothesis is that the surge in demand for online advertising is the cause for the rising prices of online advertising. While CPM is a great measure for the average expensiveness of advertising, it does not provide the complete mapping between advertising expenditure and impressions, especially when such mappings are non-linear. For example, larger, more influential advertisers may be able to secure better deals with TV networks, while smaller ones do not enjoy the same bargaining power. Therefore, it is important to look at other measures for the cost of advertising as well.

4.2.2 Marginal Cost of Advertising

To find the mapping between advertising spending and impression, I use the following reduced-form regression equation:

$$\log(\mathbb{I}_{i,t}) = \beta \log(\eta_{i,t}) + \alpha_i + \kappa_t + \epsilon_{i,t} \tag{25}$$

where $\mathbb{I}_{i,t}$ is the impressions for firm i's ads in month t, $\eta_{i,t}$ is the total advertising spending, α_i and κ_t are firm and time fixed effects, and $\epsilon_{i,t}$ is the error term. Here both impressions and advertising spendings are aggregated to the firm level, where I take the sum across different media types that provide impressions data (TV, radio and internet)[7]. I define the regression

[7] An impression is defined as exposure to either a 30-second TV commercial, a 30-second radio commercial, or an online advertisement.

coefficient β as the "impression elasticity" of advertising, which measures the percentage increase in impressions as a response to a 1% increase in advertising spendings. We can also interpret the impression elasticity as a measure of marginal cost of advertising, where higher elasticities are equivalent to lower marginal costs.

Figure 7 shows the point estimates of impression elasticity from 2010 to 2016, as well as the 95% confidence bands. We see that the impression elasticity of advertising has been steadily increasing, from 0.94 in 2010 to around 0.99 in 2016. The result shows that the relationship between advertising spending and impressions is becoming closer to linear. Whether that result is driven by the proliferation of internet advertising or by changes within other media platforms is beyond the scope of this paper. Our main argument here is that the aggregate cost structure of advertising in 2016 is different from that in 2010, regardless of the underlying reasons.

To further illustrate this point, I compare the impression elasticities in 2010 and 2016 using quantile regressions as in Koenker and Hallock (2001), and Figure 8 shows the result. We see that impression elasticities in 2016 are higher than the levels in 2010 for advertisers of all sizes, but the improvement is greatest for medium to large advertisers. Another striking difference is in the shares of advertisers with increasing marginal returns from advertising, which correspond to the sections in Figure 8 with positive slopes. In 2010, only advertisers in the bottom 25th percentile have increasing marginal returns; in 2016, this cutoff rise to 50th percentile.

In sum, the marginal cost of advertising has been decreasing from 2010 to 2016, and the distribution of marginal costs across advertisers have shifted as well.

5 Quantitative Model

To explore the role of advertising on firm sales and new product lifecycles, we construct a general equilibrium model with heterogeneous multi-product firms that make endogenous advertising decisions. The model is most similar to the theoretical framework in Hottman, Redding and Weinstein (2016), with an upper-level Cobb-Douglas demand system across product groups, nested with CES demand across firms and products.

5.1 Environment

5.1.1 Demand

Utility U_t is defined as:

$$\ln U_t = \int_{g\in\Omega_g} \varphi_{gt}^G \ln C_{gt}^G \, dg, \quad \int_{g\in\Omega^G} \varphi_{gt}^G dg = 1 \tag{26}$$

where g denotes a product group, φ_{gt}^G the expenditure share on product group g at time t, and Ω_g the set of all product groups. In addition, two CES nests for firms and UPCs can be written as:

$$C_{gt}^G = \left[\sum_{f\in\Omega_{gt}^F} \left(\varphi_{fgt}^F C_{fgt}^F\right)^{\frac{\sigma_g^F-1}{\sigma_g^F}}\right]^{\frac{\sigma_g^F}{\sigma_g^F-1}} \quad , \quad C_{fgt}^F = \left[\sum_{u\in\Omega_{fgt}^U} \left(\varphi_{ut}^U C_{ut}^U\right)^{\frac{\sigma_g^U-1}{\sigma_g^U}}\right]^{\frac{\sigma_g^U}{\sigma_g^U-1}} \tag{27}$$

In other words, consumption in each product group C_{gt}^G is a function of firm output C_{fgt}^F, which in turn is a function of consumption of each UPC, denoted by C_{ut}^U. The CES weights φ_{ut}^U and φ_{fgt}^F represent consumer appeal of each UPC and firm, defined as utility per unit of consumption[8]. Because the utility function is homogeneous with degree zero on both φ_{fgt} and φ_{ut}, the following normalization is necessary:

$$\tilde{\varphi}_{gt}^F = \left(\prod_{f\in\Omega_{gt}^F} \varphi_{fgt}^F\right)^{\frac{1}{N_{gt}^F}} = 1 \quad , \quad \tilde{\varphi}_{fgt}^F = \left(\prod_{u\in\Omega_{fgt}^U} \varphi_{ut}^U\right)^{\frac{1}{N_{fgt}^U}} = 1 \tag{28}$$

Where N_{gt}^F is the number of firms in product group g at time t, and N_{fgt}^U the number of products (UPCs) produced by firm f in product group g at time t.

[8]Differences in product and firm appeals may arise from variations in product quality or consumer taste. In the theoretical model, we stay agnostic towards different interpretations of product appeals.

For consumptions defined in equation (27), the corresponding exact price indexes are:

$$P_{gt}^{G} = \left[\sum_{f \in \Omega_{gt}^{F}} \left(\frac{P_{fgt}^{F}}{\varphi_{fgt}^{F}} \right)^{1-\sigma_g^F} \right]^{\frac{1}{1-\sigma_g^F}} \quad , \quad P_{fgt}^{F} = \left[\sum_{u \in \Omega_{fgt}^{U}} \left(\frac{P_{ut}^{U}}{\varphi_{ut}^{U}} \right)^{1-\sigma_g^U} \right]^{\frac{1}{1-\sigma_g^U}} \tag{29}$$

Different from traditional price indexes in Dixit-Stiglitz demand systems, these indexes are calculated using prices adjusted by product and firm appeals. The parameters φ_{fgt} and φ_{ut} capture changes in consumer tastes over time, for individual goods and the distribution across all goods[9].

5.1.2 Technology

To capture heterogeneity in firm productivity, I allow cost functions to vary across products and firms. Firms pay both variable and fixed costs to operate in the market. The variable cost function for product u at time t is

$$\Theta_{ut}(Y_{ut}^{U}) = \theta_{ut}(Y_{ut}^{U})^{1+\delta_g} \tag{30}$$

where θ_{ut} is a cost shifter. Firms also pay a fixed cost H_{gt}^{F} to enter a product group and H_{gt}^{U} for each unique variety sold in that product group. In addition, a firm can spend η_{fgt} on its advertising, which affects its "brand preferences" relative to other firms in the same product group:

$$\log \varphi_{fgt}^{F} = \rho_{gt} \left(\log q(\eta_{fgt}^{F}) - \frac{1}{N_{gt}^{F}} \sum_{f' \in \Omega_{gt}^{F}} \log q(\eta_{f'gt}^{F}) \right) + (1-\rho_{gt}) \mathbb{G}_{fgt}^{F} \tag{31}$$

Similar to the one-sector model, q is the advertising impression function, and it satisfies $q(0) > 0$, $\lim_{x\to\infty} q(x) \leq \infty$, $q'(x) > 0$, $q''(x) < 0$ for all $x \in [0, \infty)$. But different from the one-sector model, I add another coefficient ρ_g to represent the share of consumer brand preferences determined by current period advertising. The remaining $(1 - \rho_{gt})$ is determined

[9] See Redding and Weinstein (2019) for a comprehensive discussion on how to calculate and measure aggregate price indexes with consumer taste shocks.

by "goodwill" of a brand, denoted by $\mathbb{G}^F_{fgt}$, that can depend on previous period advertising levels, previous period sales, as well as other unobservable factors that influence demands, such as product placement and packaging. I assume that firms treat the goodwill of their brands as given at the beginning of each period, and do not take into account the impact of current period advertising on future goodwill and demands.

5.1.3 Profit Maximization

Each firm f in product group g choose its set of products $u \in \{\underline{u}_{fgt}, \ldots, \bar{u}_{fgt}\}$, prices $\{P^U_{ut}\}$ and advertising expenditure η^F_{fgt}, taking into account of its influence on aggregate price indexes:

$$\max_{\{\underline{u}_{fgt},\ldots,\bar{u}_{fgt}\},\{P^U_{ut}\},\{\eta^F_{fgt}\}} \sum_{k=\underline{u}_{fgt}}^{\bar{u}_{fgt}} [P^U_{kt}Y^U_{kt} - \Theta^U_{kt}(Y^U_{kt})] - N^U_{fgt}H^U_{fgt} - H^F_{gt} - \eta^F_{fgt} \quad (32)$$
$$\text{s.t. } Y^U_{kt} = C^U_{kt}$$

One feature of this framework is that in equilibrium, markups across products within the same firm are the same, at each given time t. In other words, markups only vary at the firm level[10]:

$$\mu^F_{fgt} \equiv \frac{P_{ut}}{\gamma_{ut}} = \frac{\epsilon^F_{fgt}}{\epsilon^F_{fgt} - 1} \quad (33)$$

Here γ_{ut} is the marginal cost to produce good u, and ϵ^F_{fgt} is the firm's perceived elasticity of demand, defined as:

$$\epsilon^F_{fgt} = \sigma^F(1 - S^F_{fgt}) + S^F_{fgt} \quad (34)$$

We can also solve the revenue share of firm f in product group g as well as the revenue share

[10]This result was proven by Hottman, Redding and Weinstein (2016), in Appendix A of their paper.

of product u in firm f:

$$S_{fgt}^{F}=\frac{\left(\frac{P_{fgt}^{F}}{\varphi_{fgt}^{F}}\right)^{1-\sigma^{F}}}{\sum_{k\in\Omega_{gt}^{F}}\left(\frac{P_{kgt}^{F}}{\varphi_{kgt}^{F}}\right)^{1-\sigma^{F}}}\quad,\quad S_{ut}^{U}=\frac{\left(\frac{P_{ut}^{U}}{\varphi_{ut}^{U}}\right)^{1-\sigma^{U}}}{\sum_{k\in\Omega_{fgt}^{U}}\left(\frac{P_{kt}^{U}}{\varphi_{kt}^{U}}\right)^{1-\sigma^{U}}} \tag{35}$$

Finally, the demand for each UPC is:

$$C_{ut}^{U}=(\varphi_{fgt}^{F})^{\sigma^{F}-1}(\varphi_{ut}^{U})^{\sigma^{U}-1}E_{gt}^{G}(P_{gt}^{G})^{\sigma^{F}-1}(P_{fgt}^{F})^{\sigma^{U}-\sigma^{F}}(P_{ut}^{U})^{-\sigma^{U}} \tag{36}$$

where E_{gt}^{G} denotes the total sales in product group g in time t.

5.1.4 Optimal Level of Advertising

The optimal amount of advertising can either be positive or zero. From the Kuhn-Tucker conditions in (32), the following relationship need to hold when optimal advertising expenditure is positive:

$$\sum_{k=\underline{u}_{fgt}}^{\bar{u}_{fgt}}(P_{kt}^{U}-\gamma_{ut})\frac{\partial Y_{kt}^{U}}{\partial\eta_{fgt}}=1,\quad \text{if}\ \ \eta_{fgt}^{*}>0 \tag{37}$$

If optimal advertising level η_{fgt}^{*} is equal to 0, then the left hand side of (37) is less than or equal to 1. Marginal cost γ_{ut} is equal to

$$\gamma_{ut}\equiv\Theta_{ut}'(Y_{ut})=(1+\delta_{g})\theta_{ut}(Y_{ut})^{\delta_{g}} \tag{38}$$

Note that when solving for (37), we make an implicit assumption that the number of a firm's products does not change with its advertising expenditure, i.e. $\frac{\partial N_{fgt}}{\partial\eta_{fgt}}=0$. This assumption will be dropped later when we study the effect of advertising on new product entry.

Using the equilibrium pricing rule in (33), we can rewrite the Kuhn-Tucker condition in (37) as:

$$\sum_{k=\underline{u}_{fgt}}^{\bar{u}_{fgt}}P_{kt}^{U}\frac{\partial Y_{kt}^{U}}{\partial\eta_{fgt}}=\epsilon_{fgt}^{F},\quad \text{if}\ \ \eta_{fgt}^{*}>0 \tag{39}$$

The left hand side of (39) is the "marginal value of advertising", defined as the firm's revenue gain from a marginal increase in advertising spending, holding all prices constant. The right hand side is the firm's perceived elasticity of demand. Intuitively, a firm can either cut prices or buy ads to boost its sales. The marginal revenue gain from these two competing methods must be equal, if spending any positive amount on advertising is optimal. This result is a generalization of the findings in Dorfman and Steiner (1954), but with multi-product firms instead.

5.1.5 Sales Effect

In this section, we analyze the sales effect of advertising for each firm, assuming the total number of its products stays unchanged. From the UPC demand function in (36), the sales effect of advertising on each UPC is:

$$\frac{\partial Y_{ut}^U}{\partial \eta_{fgt}} = (\sigma^F - 1)\frac{Y_{ut}^U}{\varphi_{fgt}^F}\frac{\partial \varphi_{fgt}^F}{\partial \eta_{fgt}} + (\sigma^F - 1)\frac{Y_{ut}^U}{P_{gt}^G}\frac{\partial P_{gt}^G}{\partial \eta_{fgt}} \tag{40}$$

The first term on the right hand side is the direct sales effect of advertising from higher firm appeals, while the second term is the indirect sales effect from a lower product-group price index[11]. As shown in Appendix A.2, we can simplify equation (40) further into:

$$\frac{\partial Y_{ut}^U}{\partial \eta_{fgt}} = (\sigma^F - 1)Y_{ut}^U\left[\frac{N_{gt}^F - 1}{N_{gt}^F}\frac{q'(\eta_{fgt})}{q(\eta_{fgt})}\right](1 - S_{fgt}^F) \tag{41}$$

Note that each product group's market structure has an influence on the sales effect of advertising. For example, suppose a firm is the monopoly in its product group ($N_{gt}^F = 1, S_{fgt}^F = 1$). In this case, the sales effect of advertising is 0 for all products sold by the monopoly, because the firm has no incentive to increase its appeal relative to other firms (there are none). On the contrary, suppose the market structure of a product group is competitive ($N_{gt}^F \to \infty, S_{fgt}^F \to 0$). Then the indirect effect of advertising is close to 0, as each individual firm's advertising decisions has virtually no effect on product group price

[11]For example, as we can see from the definition of P_{gt}^G in (29), if we double all firm appeal terms φ_{fgt}^F , the resulting product-group price index is halfed.

index P^G_{gt}.

We can also solve for a decision rule of each firm's advertising expenditure. Plug (41) into the Kuhn-Tucker condition in (39), we can find the relationship between a firm's sales share S^F_{fgt} and its optimal level of advertising η^*_{fgt}:

$$\begin{aligned} \frac{q'(\eta^*_{fgt})}{q(\eta^*_{fgt})} &= \frac{\sigma^F(1-S^F_{fgt})+S^F_{fgt}}{(\sigma^F-1)(1-S^F_{fgt})S^F_{fgt}E^G_{gt}} \cdot \frac{N^F_{gt}}{N^F_{gt}-1}, \quad \text{if } \eta^*_{fgt} > 0 \\ \frac{q'(0)}{q(0)} &< \frac{\sigma^F(1-S^F_{fgt})+S^F_{fgt}}{(\sigma^F-1)(1-S^F_{fgt})S^F_{fgt}E^G_{gt}} \cdot \frac{N^F_{gt}}{N^F_{gt}-1}, \quad \text{if } \eta^*_{fgt} = 0 \end{aligned} \tag{42}$$

Figure 9 illustrates this firm decision rule. The U-shaped curves are the right hand side of (42) as a function of firm's market share S^F_{fgt}, under different values of σ^F. As the graph shows, firms with market shares closer to 0 or 1 do not advertise. The intuition is as follows. First, we know from (34) that firms with tiny market shares ($S^F_{fgt} \approx 0$) face higher elasticity of demand ($\epsilon^F_{fgt} \approx \sigma^F$) from their customers. Therefore, these tiny firms do not choose to advertise because they could instead cut prices and attract more sales[12]. Second, firms with large market shares ($S_{fgt} \approx 1$) do not advertise either, because the marginal return from advertising is smaller when the firm's sales share is closer to 1. Imagine a firm that owns 99% of the market share in its product group. This firm is not likely to spend heavily in advertising just to compete for the remaining 1% of market share.

Another implication of the model is that in product groups with higher cross-firm elasticity of substitution σ^F, a greater share of firms participate in advertising. With larger σ^F, products across firms are closer substitutes, so an incremental increase in a firm's appeal brings significant revenue and profit growth. As σ^F approaches infinity, the right hand side of (42) converge to $\frac{1}{S^F_{fgt}E^G_{gt}}\frac{N^F_{gt}}{N^F_{gt}-1}$ in the limit. This means all firms with market shares above a threshold $\tilde{S}^F_{fgt}$ choose to advertise in equilibrium:

$$\tilde{S}^F_{fgt} \equiv \frac{q(0)}{q'(0)E^G_{gt}}\frac{N^F_{gt}}{N^F_{fgt}-1}$$

[12]In other words, the small firms can cut their prices without causing large impacts on the product-group price indexes. This is why they have higher elasticity of demand than firms with larger market shares.

Figure 10 compares the decision rules under different model parameters, holding other variables fixed. The share of advertisers as a percentage to all firms is higher when product group is larger (left panel) or more competitive (right panel).

5.1.6 Product Entry Effect

In previous sections, we focused on the sales effect of advertising at each UPC and firm level, assuming that total number of products is fixed. In equilibrium, the number of products supplied by each firm f within product group g, N^U_{fgt}, is endogenously determined by the zero profit condition. This condition requires that a firm's total profit from selling $N^U_{fgt}+1$ products is no greater than its profits from N^U_{fgt} products. Formally, the zero profit condition is:

$$\sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}+1} \pi^U_{ut}(N^U_{fgt}+1) - \sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}} \pi^U_{ut}(N^U_{fgt}) \leq H^U_{gt} \tag{43}$$

Where $\pi^U_{ut}(N^U_{fgt})$ is the variable profit function for UPC u when firm f supplies N^U_{fgt} types of products within product group g. In equilibrium, the profit function can be written as:

$$\pi^U_{ut}(N^U_{fgt}) = P^U_{ut}Y^U_{ut} - \Theta_{ut}(Y^U_{ut}) = \left(\frac{(1+\delta_g)\mu^F_{fgt}-1}{(1+\delta_g)\mu^F_{fgt}}\right) P^U_{ut}Y^U_{ut} \tag{44}$$

where δ_g is the elasticity of marginal costs with respect to output, and μ^F_{fgt} is the firm markup as defined in (33). Use UPC demand in (36), rewrite the profit function as:

$$\pi^U_{ut}(N^U_{fgt}) = \kappa \left[E^G_{gt}(\varphi^F_{fgt})^{\sigma^F-1}(P^G_{gt})^{\sigma^F-1}(P^F_{fgt})^{\sigma^U-\sigma^F}\right]\left(\frac{P^U_{ut}}{\varphi^U_{ut}}\right)^{1-\sigma^U} \tag{45}$$

where $\kappa \equiv \frac{(1+\delta_g)\mu^F_{fgt}-1}{(1+\delta_g)\mu^F_{fgt}}$. Sum over u and use the definition of firm price index in (29) to solve the firm profit function:

$$\sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}} \pi^U_{ut}(N^U_{fgt}) = \kappa E^G_{gt}(\varphi^F_{fgt})^{\sigma^F-1}(P^G_{gt})^{\sigma^F-1}(P^F_{fgt})^{1-\sigma^F} \tag{46}$$

The notations in (46) need some clarification, as N^U_{fgt} seems to only appear at the left hand side of the equation. When a firm introduces a new good, it causes both direct and indirect effect on the firm's profit. The direct effect is through changes in the firm's price index, P^F_{fgt}. The indirect effect is when the updated firm price index further affects product group price index P^G_{gt}, market share S^F_{fgt} and markup μ^F_{fgt}. If the market is competitive, the indirect effect will be small, because each firm's price levels hardly affect its market share and other product-group level variables. In this case, we can rewrite the left hand side of zero profit condition in (43):

$$\begin{aligned}&\sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}+1} \pi^U_{ut}(N^U_{fgt}+1) - \sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}} \pi^U_{ut}(N^U_{fgt}) \\ =\;& \kappa E^G_{gt}(\varphi^F_{fgt})^{\sigma^F-1}(P^G_{gt})^{\sigma^F-1}\left[[P^F_{fgt}(N^U_{fgt})]^{1-\sigma^F} - [P^F_{fgt}(N^U_{fgt}+1)]^{1-\sigma^F}\right] \qquad (47)\end{aligned}$$

where $P^F_{fgt}(N^U_{fgt})$ is the firm price index when it supplies N^U_{fgt} unique varieties of products. Note that we assume the product group price index P^G_{gt} and firm markup μ^F_{fgt} are unchanged from entry of the new product.

The profit difference in (47) is increasing in advertising expenditure through higher firm appeal, φ^F_{fgt}. In other words, a firm's profit gain from introducing a new product is higher when the firm spends more on advertising relative to other firms. The model therefore implies that the equilibrium number of product per firm, N^F_{fgt}, is positively correlated with the firm's advertising expenditure.

Consider a special case when firms are monopolistic competitors ($S^F_{fgt} \approx 0$) and all goods are equally substitutable within firms and across firms ($\sigma^U = \sigma^F$). The profit difference in (47) then becomes:

$$\begin{aligned}\sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}+1} \pi^U_{ut}(N_{fgt}+1) - \sum_{u=\underline{u}_{fgt}}^{\bar{u}_{fgt}} \pi^U_{ut}(N_{fgt}) &= \kappa E^G_{gt}(\varphi^F_{fgt})^{\sigma^F-1}(P^G_{gt})^{\sigma^F-1}\left(\frac{P^U_{\bar{u}t}}{\varphi^U_{\bar{u}t}}\right)^{1-\sigma^U} \\ &= \pi^U_{\bar{u}t}(N_{fgt}+1)\end{aligned}$$

where $P^U_{\bar{u}t}$, $\varphi^U_{\bar{u}t}$ and $\pi^U_{\bar{u}t}$ are price, product appeal and profit of the new UPC. In this case, profits of a firm's new products have no impact on its existing products. As discussed in Hottman, Redding and Weinstein (2016), this is when the "cannibalization effect" from new products is 0. The decision of whether to introduce a new product depends entirely on whether the expected profit collected from the new product would exceed the fixed entry cost. When a firm spends more aggressively in advertising, profits from its new product increases, allowing the firm to introduce more product varieties.

5.2 Structural Estimation

Our structural estimation of the model takes the following three steps. First, we estimate the model parameters $\{\sigma_g^U, \sigma_g^F, \delta_g\}$ using the same technique as in Feenstra (1994), Broda and Weinstein (2006, 2010) and Hottman, Redding and Weinstein (2016). Second, using estimated values of $\{\sigma_g^U, \sigma_g^F, \delta_g\}$, we use the model to calculate values of $\{\varphi_{fgt}^F, \varphi_{ut}^U, \theta_{ut}\}$ up to a normalization. Finally, we use the firm appeals φ_{fgt}^F and advertising expenditure data to estimate the shape of $q(\cdot)$ from equation (31).

5.2.1 UPC Moment Conditions

To estimate the elasticity parameters $\{\sigma_g^U, \delta_g\}$, we construct a set of moment conditions by first double differencing the UPC demand shares in equation (35) over time and with respect to the largest UPC within each firm:

$$\Delta^{\underline{u},t} \ln S_{ut}^U = (1 - \sigma_g^U)\Delta^{\underline{u},t} \ln P_{ut}^U + \omega_{ut} \tag{48}$$

where u is a UPC and $\underline{u}$ is the largest UPC from the same firm that produced u. The double difference operator is defined as $\Delta^{\underline{u},t} x_{ut} = \Delta^t x_{ut} - \Delta^t x_{\underline{u}t}$. The error term is defined as $\omega_{ut} = (\sigma_g^U - 1)[\Delta^t \ln \varphi_{ut}^U - \varphi \ln \varphi_{\underline{u}t}^U]$. We also construct an equation from UPC supply, using the production technology in (30) and the optimal pricing rule in (33):

$$\Delta^{\underline{u},t} \ln P_{ut}^U = \frac{\delta_g}{1+\delta_g}\Delta^{\underline{u},t} \ln S_{ut}^U + \kappa_{ut} \tag{49}$$

where $\kappa_{ut} = \frac{1}{1+\delta_g}[\Delta^t \ln a_{ut} - \Delta^t \ln a_{\underline{u}t}]$ is the stochastic error term. Finally, we define the set of UPC moment conditions from orthogonality of double-differenced demand and supply shocks:

$$G(\zeta_g) = \mathbf{E_T}[\omega_{ut}(\zeta_g)\kappa_{ut}(\zeta_g)] = 0 \tag{50}$$

where $\zeta_g = \begin{pmatrix} \sigma_g^U \\ \delta_g \end{pmatrix}$ and $\mathbf{E_T}$ is the expectation over time. The parameters ζ_g within each product group g can be estimated using the following GMM objective function:

$$\hat{\zeta}_g = \arg\min_{\zeta_g}\{G^*(\zeta_g)'WG^*(\zeta_g)\} \tag{51}$$

where $G^*(\zeta_g)$ is constructed by stacking all the UPC moment conditions for goods in product group g. The identification of ζ_g is based on our assumption that demand and supply shocks are orthogonal, which is a standard practice in macroeconomics and international trade literature (Feenstra (1994), Broda and Weinstein (2006, 2010)). Furthermore, as discussed in Leamer (1981) and Feenstra (1994), the UPC moment conditions in (50) define a rectangular hyperbola on the (σ_g^U, δ_g) space. The hyperbolas are different for each pair of UPCs, if the double-differenced demand and supply shocks are heteroskedastic. The intersection of these hyperbolas can then be used to identify (σ_g^U, δ_g), even though we do not have instruments for demand or supply.

5.2.2 Firm Moment Conditions

To estimate the remaining parameter σ_g^F, we can construct the firm moment conditions using a similar method. More specifically, we use equation (35) and observed UPC expenditure shares (S_{ut}^U) and prices (P_{ut}^U) to determine UPC appeals (φ_{ut}^U) up to our normalization. We then use the calculated UPC appeals and their observed prices to calculate firm price indexes (P_{fgt}^F) from (29). Next, we double difference log firm shares in (35), with respect to time and also the largest firm within each product group, to obtain the following equation:

$$\Delta^{\underline{f},t} \ln S_{fgt}^F = (1 - \sigma_g^F)\Delta^{\underline{f},t} \ln P_{fgt}^F + \omega_{fgt} \tag{52}$$

where $\Delta^{\underline{f},t}$ is the double difference operator over time and relative to the largest firm $\underline{f}$ in each product group, and the error term is $\omega_{fgt} = (\sigma_g^F - 1)\Delta^{\underline{f},t} \ln \varphi_{fgt}^F$.

Equation (52) cannot be estimated using OLS, because the firm appeals in the stochastic error may be correlated with firm prices. To solve this endogeneity problem, we can rewrite the firm price index as:

$$\ln P_{fgt}^F = \ln \tilde{P}_{fgt}^U + \frac{1}{1-\sigma_g^U} \ln \left[\sum_{u \in \Omega_{ft}^U} \frac{S_{ut}^U}{\tilde{S}_{fgt}^U} \right] \tag{53}$$

where tilded variables are the geometric means across UPCs within the same firm. Here, firm price indexes can be decomposed into two terms. The first term on the right hand side is a traditional Jevons price index(a geometric mean of all product prices), and the second term captures the dispersion of UPC market shares within the firm to adjust for the price indexes in a multiproduct firm. The double differenced firm price index is therefore:

$$\Delta^{\underline{f},t} \ln P_{fgt}^F = \Delta^{\underline{f},t} \ln \tilde{P}_{fgt}^U + \frac{1}{1-\sigma_g^U} \Delta^{\underline{f},t} \ln \left[\sum_{u \in \Omega_{ft}^U} \frac{S_{ut}^U}{\tilde{S}_{fgt}^U} \right] \tag{54}$$

As in Hottman et al. (2016), we use the second term on the right hand side as an instrument for the double differenced firm price index, as it only affects firm sales share (S_{fgt}^F) through the firm price index (P_{fgt}^F). This step allows us to estimate the elasticity of substitution across firms within a product group (σ_g^F).

5.2.3 Advertising Moment Conditions

So far, we have used UPC and firm moment conditions to estimate the model parameters $\{\sigma_g^F, \sigma_g^U, \delta_g\}$. We can then calculate unobserved structural residuals $\{\varphi_{ut}^U, \varphi_{fgt}^F, a_{ut}\}$ from the model, following the same steps as in Hottman et al. (2016). An important (and novel) feature of our model is that a firm's appeal φ_{fgt}^F is determined by its advertising "impact" relative to other firms in the same product group. In this section, we impose a specific functional form for the advertising impact function $q()$, and estimate its parameters to test the validity of our assumptions.

We assume the advertising impression function takes the following form:

$$q_g(\eta^F_{fgt}) = \kappa^F_{fgt}\,(1+\eta_{fgt})^{\beta_g} \tag{55}$$

where β_g is the product-group level coefficient that captures the impression elasticity of advertising, as defined in the simpler model. In addition, we assume that the multiplicative coefficient κ^F_{fgt} takes the following form:

$$\kappa^F_{fgt} = \alpha_g \cdot \alpha_t \cdot \kappa_f \cdot \epsilon_{fgt} \tag{56}$$

The first two parameters are product group and time fixed effects, respectively. The third parameter measures firm-level heterogeneity in ad effectiveness, which we assume is a random variable from a log-normal distribution. Finally, the error term ϵ_{fgt} captures the remaining differences in ad effectiveness.

Using this specification, we can take logarithms on both sides of equation (31):

$$\begin{aligned}\log\varphi^F_{fgt} =& \log q(\eta^F_{fgt}) - \frac{1}{N_{gt}}\sum_{f'\in\Omega^F_{gt}} \log q(\eta^F_{f'gt})\\ =&\beta_g\left[\log(1+\eta^F_{fgt}) - \overline{\log(1+\eta^F_{fgt})}\right] + \left(\log\kappa_f - \overline{\log\kappa_f}\right)\ldots\\ &+\left(\log\epsilon^F_{fgt} - \overline{\log\epsilon^F_{fgt}}\right)\end{aligned} \tag{57}$$

Where $\overline{x_{fgt}}$ denotes average values of x_{fgt} within the same product group and time period, while $\overline{\log\kappa_f}$ denotes the pooled average “ad effectiveness” across all firms. Note that product group and time fixed effects cancel off in the above equation.

If η^F_{fgt} is directly observable, we can use a linear model with firm fixed effects to estimate β_g for each product group g. However, we only observe firm level advertising spending $\eta^F_{ft} = \sum_{g\in G_{ft}} \eta^F_{fgt}$ in the data, where G_{ft} is the set of categories in which firm f sells its products. To solve this problem, we use within-firm sales shares to impute firm-category level advertising spending:

$$\tilde{\eta}^F_{fgt} = \frac{P^F_{fgt} C^F_{fgt}}{\sum\limits_{g' \in G_{ft}} P^F_{fg't} C^F_{fg't}} \eta^F_{ft}$$

5.2.4 Result

Table 6 presents the estimation results using $\tilde{\eta}^F_{fgt}$ as approximate measures of firm-category level advertising spending. Column 1 and 2 show the estimated advertising impact elasticities β using OLS and linear panel models, when we assume $\beta_g \equiv \beta$ is constant across product categories g. In Column 3, we relax this assumption and allow β_g to vary across categories, using a linear mixed effects model with firm and time fixed effects and product category random effects. The mean advertising impact elasticity is between 0.6-0.7 in all specifications, providing evidence for our assumption that advertising expenditure is positively correlated with brand preferences.

Table 7 reports the distribution of advertising impression elaticities across product categories. The estimated elasticities have large variations, ranging from 0.03 at the bottom 1% to 2.27 at the top 1%. This result implies that in some product categories, firms can improve their brand images more easily through advertising; while in other categories such improvements are much harder to achieve. In other words, the effectiveness of advertising varies across different types of products.

To better understand this heterogeneity in advertising effectiveness, we plot the estimated elasticities against other category-level variables. Figure 11 plots the structural estimates β_g against the elasticity of substitution σ^F_g in each category. We find that in product categories where elasticities of substitution across firms are smaller, firms can more effectively increase brand preferences through advertising. The intuition is simple: it is easier to advertise for more differentiated products (say, beer) than similar ones (like oranges). To the best of my knowledge, no previous studies have documented this fact before, and our study is the first to point out the relationship between advertising effectiveness and elasticity of substitution. Note that this result is not simply driven by the different numbers of firms or advertisers across categories, as shown in Figure 12.

5.3 Counterfactual Analysis

I now turn to the counterfactual implications of our quantitative model, and explore the distributional impact of advertising on firm market shares. More specifically, I want to answer the question that if advertising technology in year y become the same as in $\tilde{y}$, how much would the distribution of firm market shares change. In our case, $y = 2016$ and $\tilde{y} = 2010$.

5.3.1 Method

I follow four steps to generate the counterfactual distribution of firm market shares. First, I calibrate the aggregate impression function in both years, and predict the firm-level counterfactual impression of year y, if advertising technology in year y becomes the same as in year $\tilde{y}$. Second, I generate the counterfactual distribution of brand preferences assuming advertising technology in year y is the same as in year $\tilde{y}$, where I use the counterfactual impression levels calculated from the previous step. Next, I apply results from our structural model to calculate the counterfactual distribution of firm market shares with a recursive algorithm. Finally, I use the first order conditions to calculate counterfactual advertising expenditure, and loop over the previous steps until the firm market shares converge.

Step 1: Counterfactual Impressions

I first calibrate the aggregate impression function in both years using the following reduced-form regression formula:

$$\log(\mathbb{I}_{fyq}) = \beta_{0,y} + \beta_{1,y}\log(\eta_{fyq}) + \alpha_q + \epsilon_{fyq} \tag{58}$$

where η_{fyq} and $\mathbb{I}_{fyq}$ are spending and impressions of firm f's advertisements in year y and quarter q. The regression equation includes time (quarter) fixed effect to control for seasonality within a year, and allows both $\beta_{0,y}$ and $\beta_{1,y}$ to vary across years. From these regression coefficients, we can use the following formula to compute counterfactual impressions of year

y, if the advertising technology is held fixed to the same level as in year $\tilde{y}$:

$$\log(\tilde{\mathbb{I}}_{fyq}) = \beta_{0,\tilde{y}} + \beta_{1,\tilde{y}} \log(\eta_{fyq}) + \alpha_q + \epsilon_{fyq} \tag{59}$$

The residual terms ϵ_{fyq} captures unobserved heterogeneity across firms and quarters. Note that the regression formula does not include firm or product category fixed effects, because here we want to focus on changes to the *aggregate* advertising technology. The justification goes as follows. If TV stations permanently charge lower cost per view for their ads, due to competition from online advertising, then the same low cost equally affects all TV advertisers, regardless of the parent companies or industries they belong. In other words, by omitting the firm and product category fixed effects, we implicitly assume that exogenous changes to advertising technology influence all firms equally.

Step 2: Counterfactual Brand Preferences

Next, I use the following regression equation to calibrate the effect of impression on brand preferences:

$$\log\left(\Phi_{fgyq}\right) = \beta_0 + \beta_{1,gy} \log \mathcal{I}_{fgyq} + \alpha_f + \alpha_g + \alpha_{gy} + \epsilon_{fgyq} \tag{60}$$

Here Φ_{fgyq} and $\mathcal{I}_{fgyq}$ are the normalized brand preferences and impressions for firm f in product module g, year y and quarter q. To compute the normalized values, I divide the original variable by its geometric mean across all firms in the same product module and quarter. We include firm, product module and product module $\times$ year fixed effects to remove cross-sectional variations from unobserved heterogeneity, and I allow the regression coefficient $\beta_{1,gy}$ to vary across product modules and years. After estimating the regression coefficients, I construct the counterfactual levels of normalized brand preferences using the following formula, assuming the advertising technology in year y becomes the same as year $\tilde{y}$:

$$\log\left(\tilde{\Phi}_{fgyq}\right) = \beta_0 + \beta_{1,g\tilde{y}} \log \tilde{\mathcal{I}}_{fgyq} + \alpha_f + \alpha_g + \alpha_{g\tilde{y}} + \epsilon_{fgyq} \tag{61}$$

Note that in equation (61), we use the (normalized) counterfactual impression levels $\tilde{\mathcal{I}}_{fgyq}$, which we impute from the firm-level counterfactual impression $\tilde{\mathbb{I}}_{fyq}$ and each firm's revenue shares across product modules. Finally, we can recover the counterfactual brand preferences $\tilde{\varphi}_{fgyq}$ from the normalized levels $\tilde{\Phi}_{fgyq}$.

Step 3: Counterfactual Firm Market Shares

The first formula in equation (35) describes the relationship between brand preferences, price indexes and firm market shares. Since the counterfactual brand preferences are computed from Step 2, we can attempt to compute counterfactual firm market shares using the following equation:

$$\tilde{S}_{fgyq} = \frac{\left(\frac{P_{fgyq}}{\tilde{\varphi}_{fgyq}}\right)^{1-\sigma^F}}{\sum_{k\in\Omega^F_{gt}}\left(\frac{P_{kgyq}}{\tilde{\varphi}_{kgyq}}\right)^{1-\sigma^F}} \tag{62}$$

where P_{fgyq} is the actual firm level price index P^F_{fgt} in equation (29), but written in a slightly different way to stay consistent with notations in the current section. We can then calculate counterfactual markups using the model:

$$\tilde{\mu}_{fgyq} = \frac{\tilde{\boldsymbol{\epsilon}}_{fgyq}}{\tilde{\boldsymbol{\epsilon}}_{fgyq}-1}, \quad \tilde{\boldsymbol{\epsilon}}_{fgyq} = \sigma^F(1-\tilde{S}_{fgyq}) + \tilde{S}_{fgyq} \tag{63}$$

Note that the counterfactual markups calculated here are different from actual markups, which means firms want to adjust their price levels under the different advertising technology. But as firms change their price indexes to $\tilde{P}_{fgtq}$, their market shares also changes, according to (62). Consequently, firms update markups because of the new market shares, which further motivate them to change price indexes, and so on. To solve this problem, I compare the counterfactual results under two parallel situations. In the first situation, I assume that firms cannot change their price levels, and compute counterfactual market shares directly using equation (62). In the second situation, I assume that firms can change their prices but not their advertising levels. The counterfactual market shares and price indexes are jointly determined, where I update each variable recursively until they both converge. The results are shown in Figure 13 of the next section.

Step 4: Counterfactual Advertising Spending

Because advertising spending is an endogenous variable, firms may want to change their marketing budget if they realize that the advertising technology in year y has changed to that in year $\tilde{y}$. Using equations (42), I solve the counterfactual levels of advertising spending $\tilde{\eta}_{fgq}$ from the . I repeat Step 1 through Step 4 until the counterfactual firm market shares, price indexes and advertising expenditures all converge. Finally, I use the converged firm market shares $\tilde{S}_{fgyq}$ to compute counterfactual markups, from equation (63). The counterfactual Herfindahl indexes are simply:

$$\tilde{\mathbb{H}}_{gyq} = \sum_{f \in \Omega_{gt}^F} (\tilde{S}_{fgyq})^2 \tag{64}$$

5.3.2 Result

Figure 13 shows the counterfactual results under three different scenarios. In all three cases, we assume advertising technologies are fixed at the 2010 level, and compute the counterfactual markups and Herfindahls under that assumption. The difference between the three cases is whether firms can freely adjust prices and advertising spendings.

We see that the counterfactual changes in both Herfindahl and markup are close to actual changes when firms cannot adjust prices or advertisement expenditures (blue bars). However, when firms *can* adjust prices but not advertisements, our counterfactual analysis suggest that both Herfindahl and markups should have been increasing from 2010 to 2016, instead of decreasing (orange bar). Finally, when firms can adjust both prices and advertising levels, both Herfindahl and markup would increase as well, while the change in Herfindahl is especially large.

6 Conclusion

In this paper, I construct a general equilibrium model with heterogeneous multi-product firms that make endogenous decisions on advertising and product entry. The model implies a cutoff rule that determines whether a firm advertises or not, and finds a U-shaped relationship between a firm's market share and its likeliness to advertise. This model also show channels where advertising positively affects firm sales and product entry. Using a novel panel data merged from Nielsen's retailer scanner data and advertising expenditure data, I find a significant but small advertising elasticity of sales, between 0.08 and 0.09. However, structural estimated results show that advertisements have large impact on firms' perceived quality, or "consumer appeals". We quantitatively estimate the magnitude of this effect, and show that the elasticity of advertising on "firm appeals" are around 0.7 on average, with large variations across product categories.

Appendix

A Mathematical Appendix

A.1 Symmetric Equilibrium in Bertrand Competition

Proof of Proposition 1

Proof. The household's problem is given by

$$\max_{\{c_i\}_{i=1}^N} \left[\sum_{i=1}^N \left(\frac{\varphi_i}{\tilde{\varphi}} c_i\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{65}$$

$$\text{s.t.} \sum_{i=1}^N p_i c_i = 1 \tag{66}$$

The Lagrangian of this problem is

$$\mathcal{L} = \left[\sum_{i=1}^N \left(\frac{\varphi_i}{\tilde{\varphi}} c_i\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} + \lambda(1 - \sum_{i=1}^N p_i c_i)$$

The first order conditions are:

$$[c_i] \quad \left(\frac{\sigma}{\sigma-1}\right) \left[\sum_{j=1}^N \left(\frac{\varphi_j}{\tilde{\varphi}} c_j\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} \left(\frac{\sigma-1}{\sigma}\right) \left(\frac{\varphi_i}{\tilde{\varphi}} c_i\right)^{-\frac{1}{\sigma}} \left(\frac{\varphi_i}{\tilde{\varphi}}\right) = \lambda p_i \tag{67}$$

$$[\lambda] \quad 1 - \sum_{i=1}^N p_i c_i = 0 \tag{68}$$

From the first order condition of any two products i and k:

$$\frac{c_k}{c_i} = \left(\frac{p_i}{p_k}\right)^{\sigma} \left(\frac{\varphi_k}{\varphi_i}\right)^{\sigma-1}$$

Use the budget constraint as well as the relationship between brand preferences and advertising in equation (2), the household's demand for product i is:

$$c_i(\boldsymbol{p}, \boldsymbol{\eta}) = \frac{p_i^{-\sigma} q(\eta_i)^{\sigma-1}}{\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}}$$

■

Proof of Proposition 2

Proof. Firm i's profit maximization problem is given by:

$$\max_{p_i, \eta_i} \; (p_i - \theta) c_i(\boldsymbol{p}, \boldsymbol{\eta}) - \eta_i$$
$$\text{s.t.} \quad \eta_i \geq 0$$

Let's first focus on the interior solutions, where $\eta_i^* > 0$ for $i = 1, 2$. The first order conditions for firm's problem is:

$$[p_i] \quad c_i(\boldsymbol{p}, \boldsymbol{\eta}) + (p_i - \theta) \frac{\partial c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial p_i} = 0 \tag{69}$$

$$[\eta_i] \quad \frac{\partial c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial \eta_i} (p_i - \theta) = 1 \tag{70}$$

Lemma 1 *(Dorfman and Steiner) In an equilibrium with positive advertising spending, the marginal increase in firm i's revenue from advertising is equal to the elasticity of demand:*

$$p_i \frac{\partial c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial \eta_i} = \epsilon_{i,p}^{D} \equiv -\frac{\partial c_i(\boldsymbol{p}, \boldsymbol{\eta})}{\partial p_i} \frac{p_i}{c_i(\boldsymbol{p}, \boldsymbol{\eta})}$$

Proof of Lemma 1. From the first order conditions above, substitute $(p_i - \theta)$ from $[\eta_i]$ to $[p_i]$, and rearrange terms. ■

Lemma 2 *(Best Response) In an equilibrium with positive advertising spending, firm i's best* *response $p_i(p_{-i}, \eta_{-i})$ and $\eta_i(p_{-i}, \eta_{-i})$ are the solutions of the following implicit functions:*

$$\frac{\theta p_i^{1-\sigma} q(\eta_i)^{\sigma-1}}{p_i(\sigma-1)-\sigma\theta} - \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1} = 0$$

$$\frac{(\sigma-1)p_i^{1-\sigma} q(\eta_i)^{\sigma-2} q'(\eta_i) \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1}}{\left[\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}\right]\left[p_i^{1-\sigma} q(\eta_i)^{\sigma-1} + \sigma \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1}\right]} - 1 = 0$$

Proof of Lemma 2. To show the first equation, take logarithm of the demand function and find the first order derivative with respect to p_i:

$$\begin{aligned}
\log c_i &= (-\sigma)\log p_i + (\sigma-1)\log(q(\eta_i)) - \log\left(\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}\right) \\
\Rightarrow \quad \frac{\partial \log c_i}{p_i} &= \frac{-\sigma}{p_i} - \frac{(1-\sigma)p_i^{-\sigma} q(\eta_i)^{\sigma-1}}{\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}} \\
&= \frac{-\sigma p_i^{1-\sigma} q(\eta_i)^{\sigma-1} - \sigma \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1} - (1-\sigma) p_i^{1-\sigma} q(\eta_i)^{\sigma-1}}{p_i\left(\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}\right)} \\
&= -\frac{p_i^{1-\sigma} q(\eta_i)^{\sigma-1} + \sigma \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1}}{p_i\left(\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}\right)} \qquad (71)
\end{aligned}$$

Plug into the first order condition $[p_i]$:

$$\begin{aligned}
&\frac{\partial \log c_i}{\partial p_i} = -\frac{1}{p_i-\theta} \\
\Rightarrow \quad &\frac{p_i^{1-\sigma} q(\eta_i)^{\sigma-1} + \sigma \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1}}{p_i\left(\sum_{j=1}^{N} p_j^{1-\sigma} q(\eta_j)^{\sigma-1}\right)} = \frac{1}{p_i-\theta} \\
\Rightarrow \quad &[(\sigma-1)p_i - \sigma\theta] \sum_{k\neq i} p_k^{1-\sigma} q(\eta_k)^{\sigma-1} = \theta p_i^{1-\sigma} q(\eta_i)^{\sigma-1}
\end{aligned}$$

This is the first equation that firm i's best response functions $p_i(\boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}})$ and $\eta_i(\boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}})$

need to satisfy. To prove the second equality, take the first order derivative of demand with respect to advertising:

$$\begin{aligned}
\frac{\partial c_i}{\partial \eta_i} =& \frac{(\sigma-1)p_i^{-\sigma}q(\eta_i)^{\sigma-2}q'(\eta_i)\left(\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}\right)}{\left(\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}\right)^2}\cdots \\
&-\frac{p_i^{-\sigma}q(\eta_i)^{\sigma-1}(\sigma-1)p_i^{1-\sigma}q(\eta_i)^{\sigma-2}q'(\eta_i)}{\left(\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}\right)^2} \\
=& \frac{(\sigma-1)p_i^{-\sigma}q(\eta_i)^{\sigma-2}q'(\eta_i)\sum_{k\neq i} p_k^{1-\sigma}q(\eta_k)^{\sigma-1}}{\left(\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}\right)^2}
\end{aligned}$$

Plug into the first order condition $[\eta_i]$:

$$\begin{aligned}
&\frac{\partial c_i}{\partial \eta_i} = \frac{1}{p_i-\theta} \\
\Rightarrow\ & \frac{(\sigma-1)p_i^{-\sigma}q(\eta_i)^{\sigma-2}q'(\eta_i)\sum_{k\neq i} p_k^{1-\sigma}q(\eta_k)^{\sigma-1}}{\left(\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}\right)^2} = \frac{1}{p_i-\theta}
\end{aligned}$$

Using the intermediate steps in the proof of last equality:

$$\begin{aligned}
&\frac{1}{p_i-\theta} = \frac{p_i^{1-\sigma}q(\eta_i)^{\sigma-1}+\sigma\sum_{k\neq i} p_k^{1-\sigma}q(\eta_k)^{\sigma-1}}{p_i\left(\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}\right)} \\
\Rightarrow\ & \frac{(\sigma-1)p_i^{1-\sigma}q(\eta_i)^{\sigma-2}q'(\eta_i)\sum_{k\neq i} p_k^{1-\sigma}q(\eta_k)^{\sigma-1}}{\sum_{j=1}^{N} p_j^{1-\sigma}q(\eta_j)^{\sigma-1}} = p_i^{1-\sigma}q(\eta_i)^{\sigma-1}+\sigma\sum_{k\neq i} p_k^{1-\sigma}q(\eta_k)^{\sigma-1}
\end{aligned}$$

This is the second equality that $p_i(\boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}})$ and $\eta_i(\boldsymbol{p_{-i}}, \boldsymbol{\eta_{-i}})$ need to satisfy.

■

In a symmetric equilibrium, $p_i = p$ and $\eta_i = \eta$ for all $i = 1, 2, \ldots, N$. Plug in the best response functions in Proposition 2, we have:

$$\frac{\theta p^{1-\sigma} q(\eta)^{\sigma-1}}{p(\sigma-1)-\sigma\theta} - (N-1)p^{1-\sigma}q(\eta)^{\sigma-1} = 0$$
$$\frac{(\sigma-1)p^{1-\sigma}q(\eta)^{\sigma-2}q'(\eta)(N-1)p^{1-\sigma}q(\eta)^{\sigma-1}}{Np^{1-\sigma}q(\eta)^{\sigma-1}\left(p^{1-\sigma}q(\eta)^{\sigma-1}+\sigma(N-1)p^{1-\sigma}q(\eta)^{\sigma-1}\right)} - 1 = 0$$

There are two equations and two unknowns (p^* and η^*), so we can solve this system of equations:

$$p^* = \frac{1+(N-1)\sigma}{(N-1)(\sigma-1)}\theta$$
$$\frac{q'(\eta^*)}{q(\eta^*)} = \frac{(1+(N-1)\sigma)}{(N-1)(\sigma-1)}N$$

Define $f(\eta) \equiv q'(\eta)/q(\eta)$. Because $q'(\eta) > 0$ and $q''(\eta) < 0$ for all $\eta \in [0, \infty)$, it is easy to show that $f(\eta)$ is a strictly decreasing in η, and $\lim_{\eta\to\infty} f(\eta) = 0$. Therefore, as long as $f(0) \geq \frac{(1+(N-1)\sigma)}{(N-1)(\sigma-1)}N$, we will have a unique solution for η^*, denoted as

$$\eta^* = f^{-1}\left(\frac{(1+(N-1)\sigma)}{(N-1)(\sigma-1)}N\right) \tag{72}$$

We now turn our focus to possible corner solutions in a symmetric equilibrium. More specifically, we replace the first order condition in firm's problem by a Kuhn-Tucker condition:

$$[\eta_i] \quad \eta_i\left(\frac{\partial c_i(\boldsymbol{p},\boldsymbol{\eta})}{\partial \eta_i}(p_i-\theta)-1\right) \geq 0$$

In a symmetric equilibrium with $\eta^* = 0$, the following condition must hold true:

$$\frac{\partial c_i(\boldsymbol{p},\boldsymbol{\eta})}{\partial \eta_i}(p_i-\theta) < 1 \quad \text{for } i = 1, 2, \ldots N$$

Using the intermediary steps above, it is easy to show that the condition is equivalent to

$$f(0) < \frac{(1 + (N-1)\sigma)}{(N-1)(\sigma-1)} N$$

■

A.2 Derivation of Equation (16)

From equation (31), we have:

$$\begin{aligned}
\frac{1}{\varphi^F_{fgt}} \frac{\partial \varphi^F_{fgt}}{\partial \eta^F_{fgt}} &= \frac{\partial \ln \varphi^F_{fgt}}{\partial \eta^F_{fgt}} \\
&= \frac{\partial}{\partial \eta^F_{fgt}} \left\{ \ln q(\eta^F_{fgt}) - \frac{1}{N^F_{gt}} \sum_{f \in \Omega^F_{gt}} \ln q(\eta^F_{fgt}) \right\} \\
&= \frac{q'(\eta^F_{fgt})}{q(\eta^F_{fgt})} \frac{N^F_{gt} - 1}{N^F_{gt}}
\end{aligned}$$

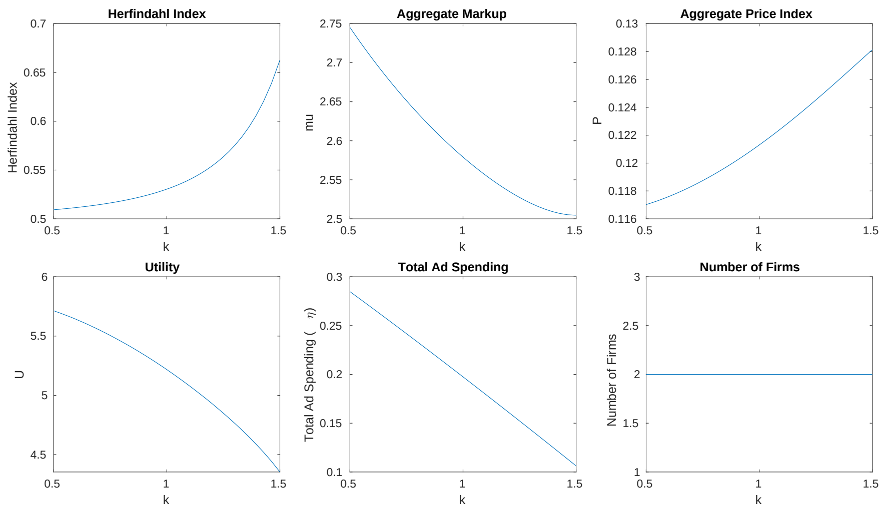
From the definition of product group price indexes in (29), we can solve:

$$\begin{aligned}
\frac{1}{P^G_{gt}} \frac{\partial P^G_{gt}}{\partial \eta^F_{fgt}} &= \frac{\partial \ln P^G_{gt}}{\partial \eta^F_{fgt}} \\
&= \frac{\partial}{\partial \eta^F_{fgt}} \left\{ \frac{1}{1-\sigma^F_g} \ln \left[\sum_{f \in \Omega^F_{gt}} \left(\frac{P^F_{fgt}}{\varphi^F_{fgt}} \right)^{1-\sigma^F_g} \right] \right\} \\
&= \frac{1}{1-\sigma^F_g} \left[\sum_{f \in \Omega^F_{gt}} \left(\frac{P^F_{fgt}}{\varphi^F_{fgt}} \right)^{1-\sigma^F_g} \right]^{-1} (\sigma^F_g - 1)(P^F_{fgt})^{1-\sigma^F_g} (\varphi^F_{fgt})^{\sigma^F_g - 2} \frac{\partial \varphi^F_{fgt}}{\partial \eta^F_{fgt}} \\
&= -S^F_{fgt} \frac{1}{\varphi^F_{fgt}} \frac{\partial \varphi^F_{fgt}}{\partial \eta^F_{fgt}}
\end{aligned}$$

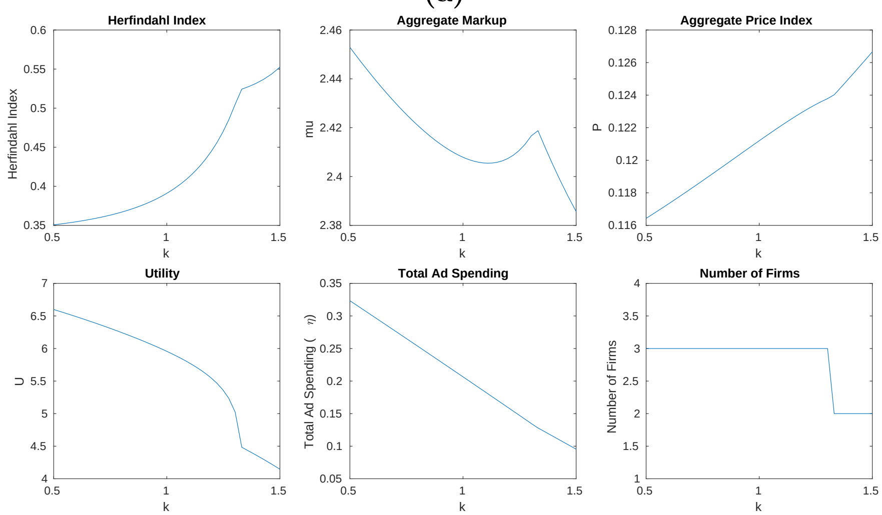
Therefore, equation (40) can be written as:

$$\begin{aligned}
\frac{\partial Y^U_{ut}}{\partial \eta^F_{fgt}} &= (\sigma^F_g - 1)\frac{Y^U_{ut}}{\varphi^F_{fgt}}\frac{\partial \varphi^F_{fgt}}{\partial \eta_{fgt}} + (\sigma^F_g - 1)\frac{Y^U_{ut}}{P^G_{gt}}\frac{\partial P^G_{gt}}{\partial \eta_{fgt}} \\
&= (\sigma^F_g - 1)Y^U_{ut}(1 - S^F_{fgt})\frac{1}{\varphi^F_{fgt}}\frac{\partial \varphi^F_{fgt}}{\partial \eta^F_{fgt}} \\
&= (\sigma^F - 1)Y^U_{ut}\left[\frac{N^F_{gt} - 1}{N^F_{gt}}\frac{q'(\eta_{fgt})}{q(\eta_{fgt})}\right](1 - S^F_{fgt})
\end{aligned}$$

B Graphs and Tables

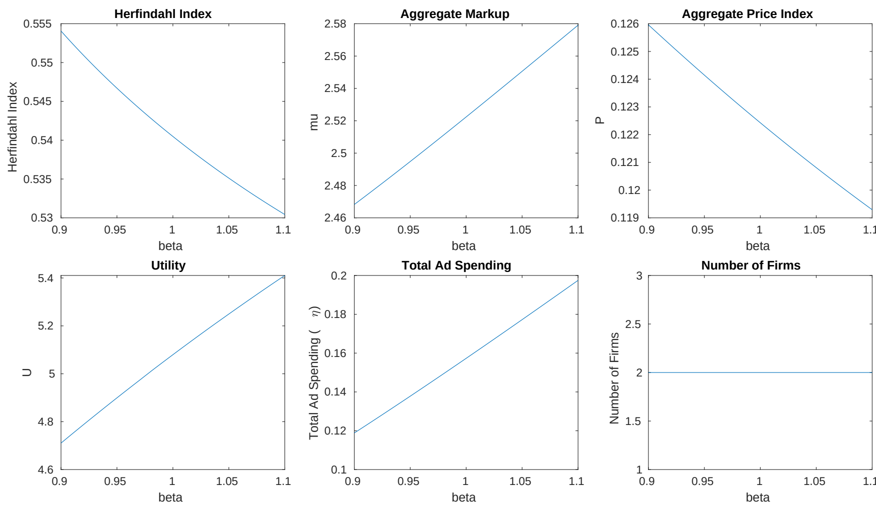


(a)

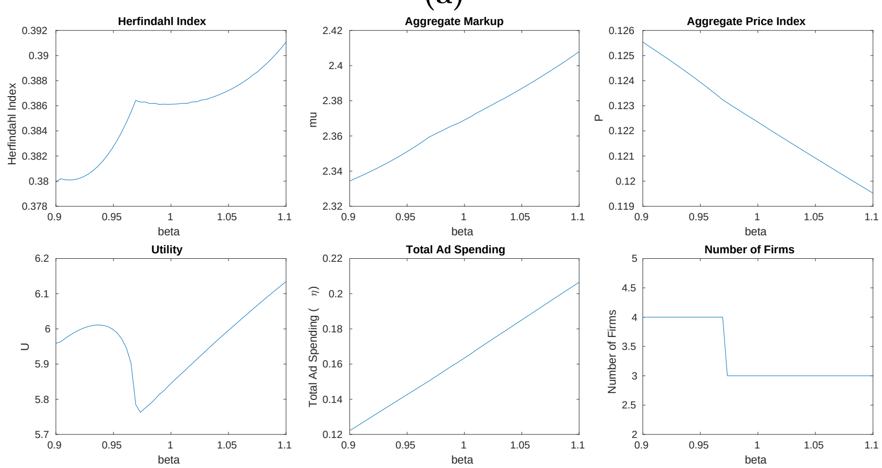


(b)

Figure 1: (a) A numerical solution for the one-sector economy under different values of K, with the parameter values $\lambda = 0.05, \beta = 1.1, \mu_\theta = 0.4$, and $\sigma_\theta = 0.1$. The economy starts with $N = 6$ firms, and I define exiting firms as the ones with market share less than 1×10^{-5}. The realized marginal costs are $\boldsymbol{\theta_A} = (0.2517, 0.4213, 0.5183, 0.5395, 0.4191, 0.3150)$. (b) Same as (a), but with marginal costs $\boldsymbol{\theta_B} = (0.3074, 0.3770, 0.2780, 0.5447, 0.3425, 0.4797)$.



(a)



(b)

Figure 2: **(a)** A numerical solution for the one-sector economy under different values of β, with the parameter values $\lambda = 0.05, K = 1, \mu_\theta = 0.4$, and $\sigma_\theta = 0.1$. The economy starts with $N = 6$ firms, and I define exiting firms as the ones with market share less than 1×10^{-5}. The realized marginal costs are $\boldsymbol{\theta_A} = (0.2517, 0.4213, 0.5183, 0.5395, 0.4191, 0.3150)$.
(b) Same as (a), but with marginal costs $\boldsymbol{\theta_B} = (0.3074, 0.3770, 0.2780, 0.5447, 0.3425, 0.4797)$.

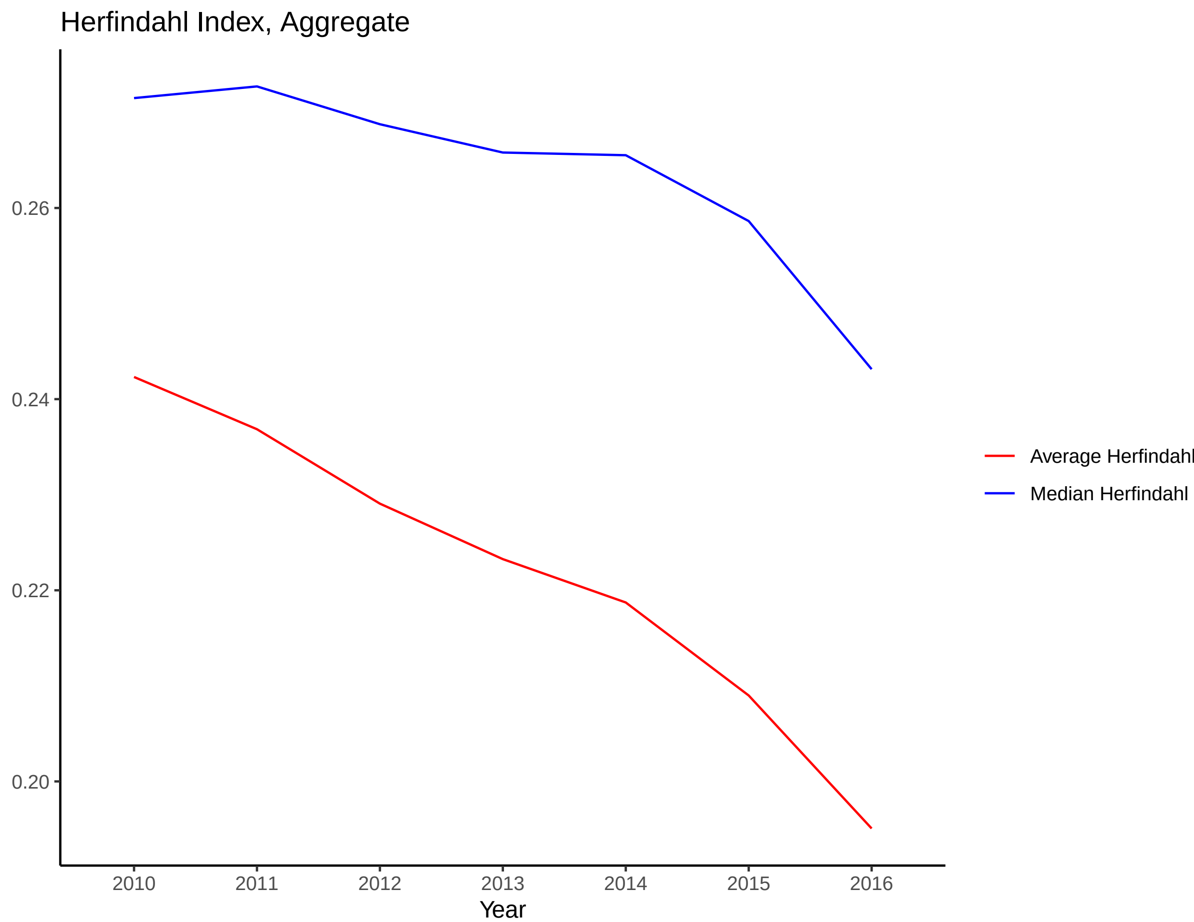


Figure 3: Mean and median Herfindahl index from 2010 to 2016, computed using quarterly firm-level market shares in each "product module", which is a narrower definition of product categories. The mean Herfindahl is the weighted average across product-module-level Herfindahls, where I use sales revenue as weights.

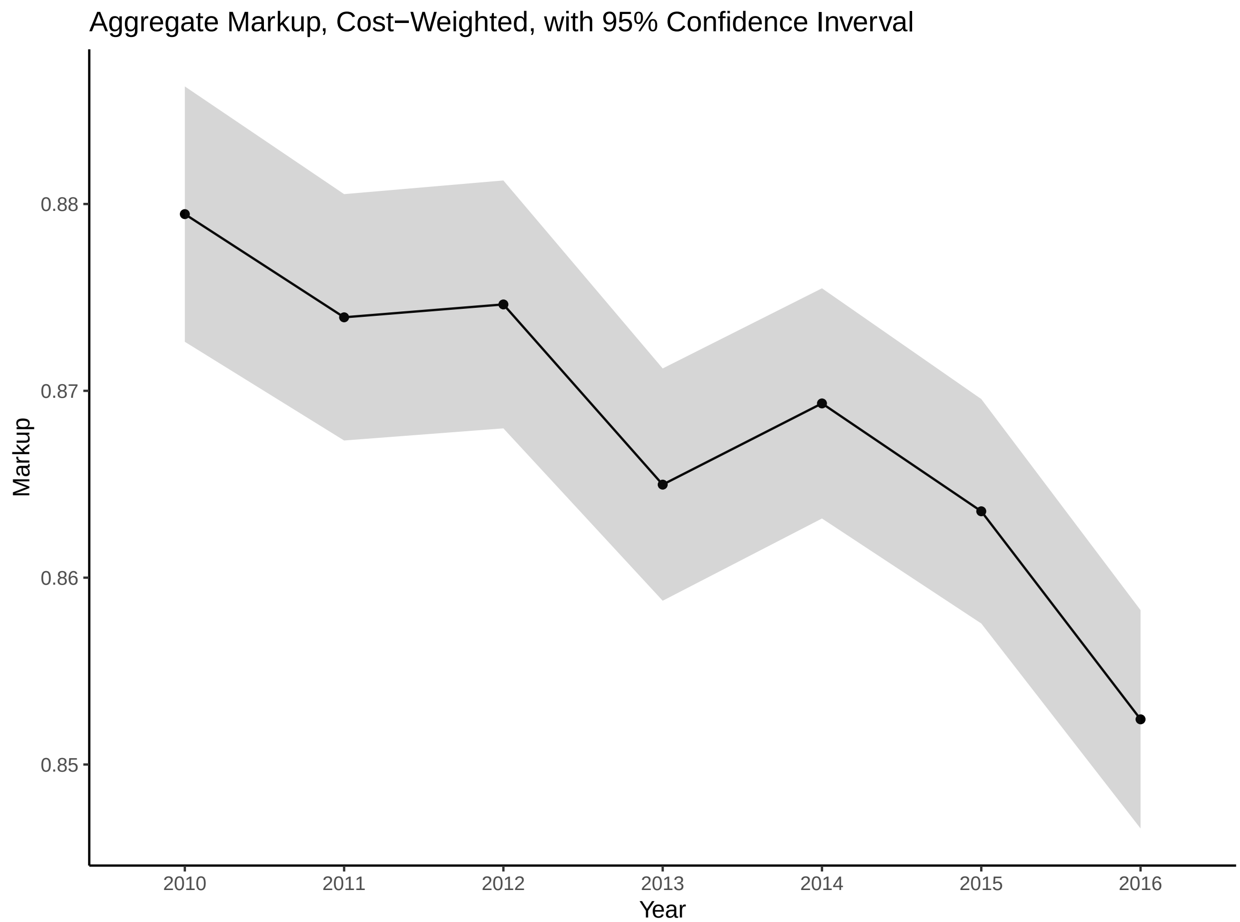


Figure 4: Aggregate markup with 95% confidence intervals, 2010-2016. Because firm markup is a nonlinear function of elasiticity of substitution, I use the Delta Method to compute the confidence bands, from the standard error of the elasticity of substitution estimated from data.

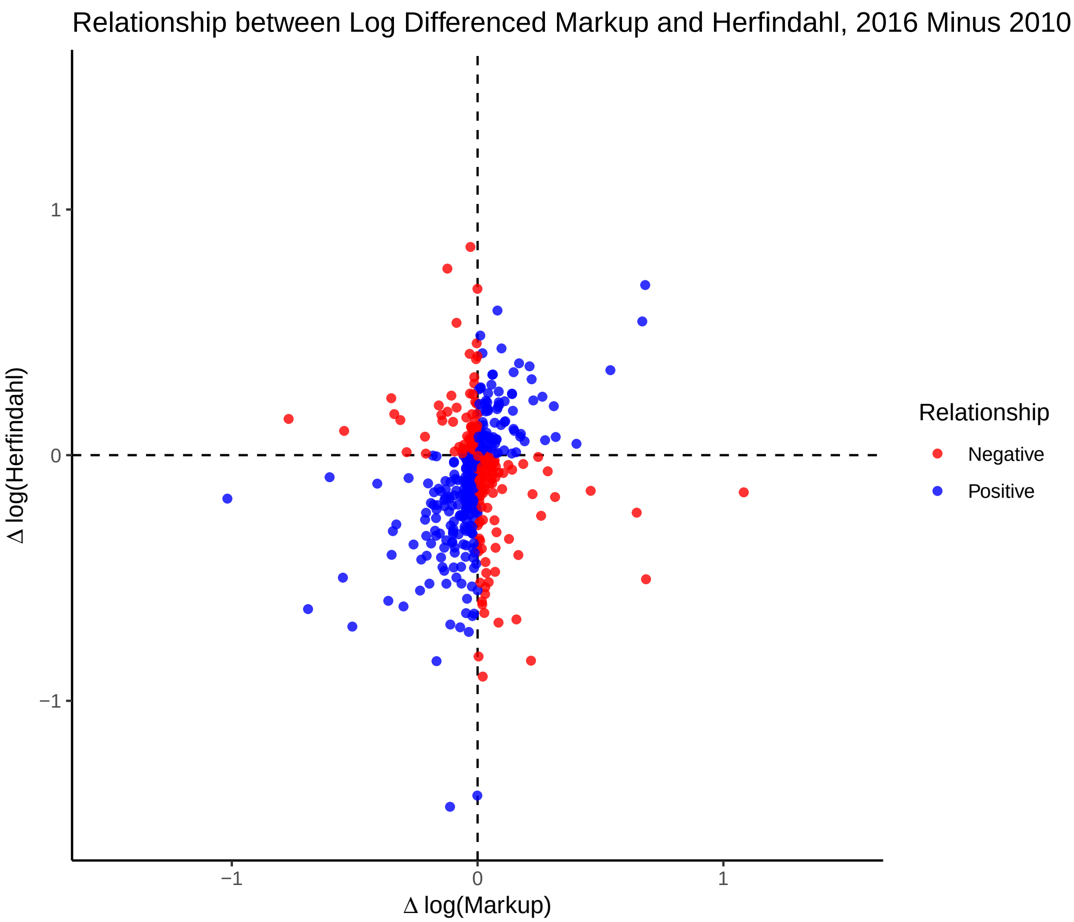


Figure 5: Changes in log markup and Herfindahl between 2010 and 2016, plotted at the product module level.

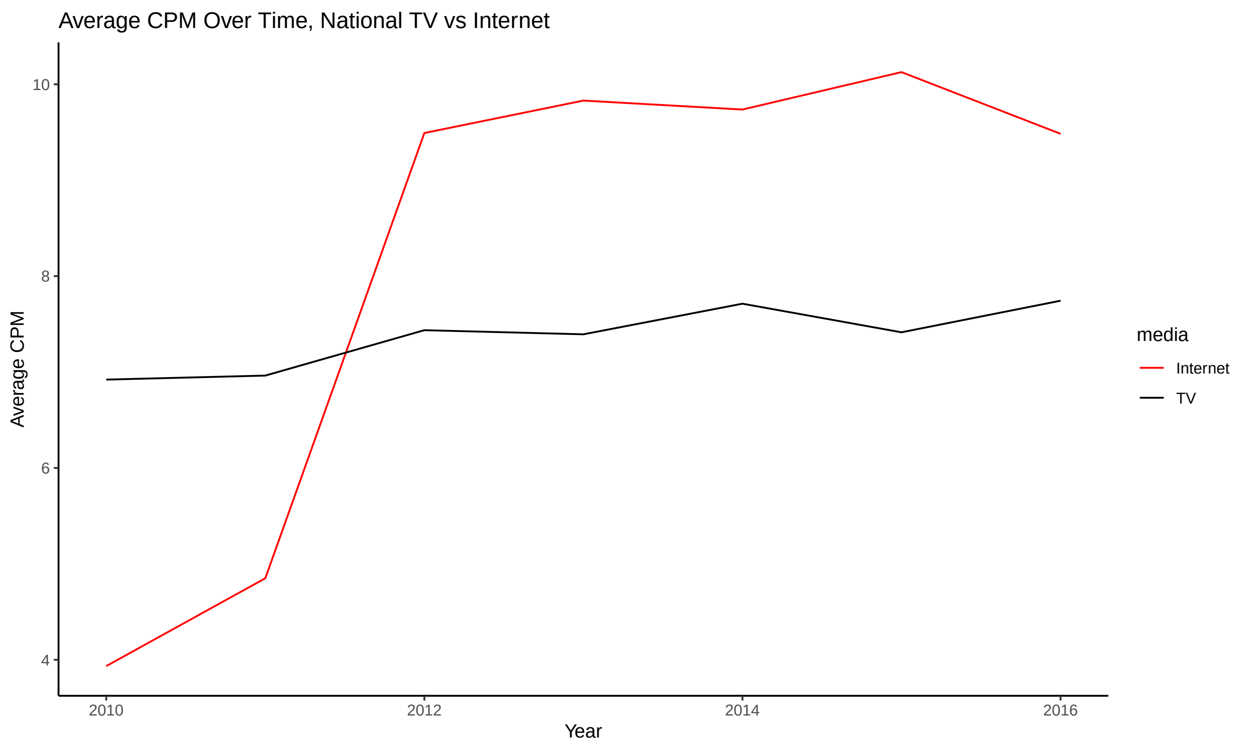


Figure 6: Average costs for national TV and internet advertisements, measured by cost per thousand views (CPM), for 2010-2016.

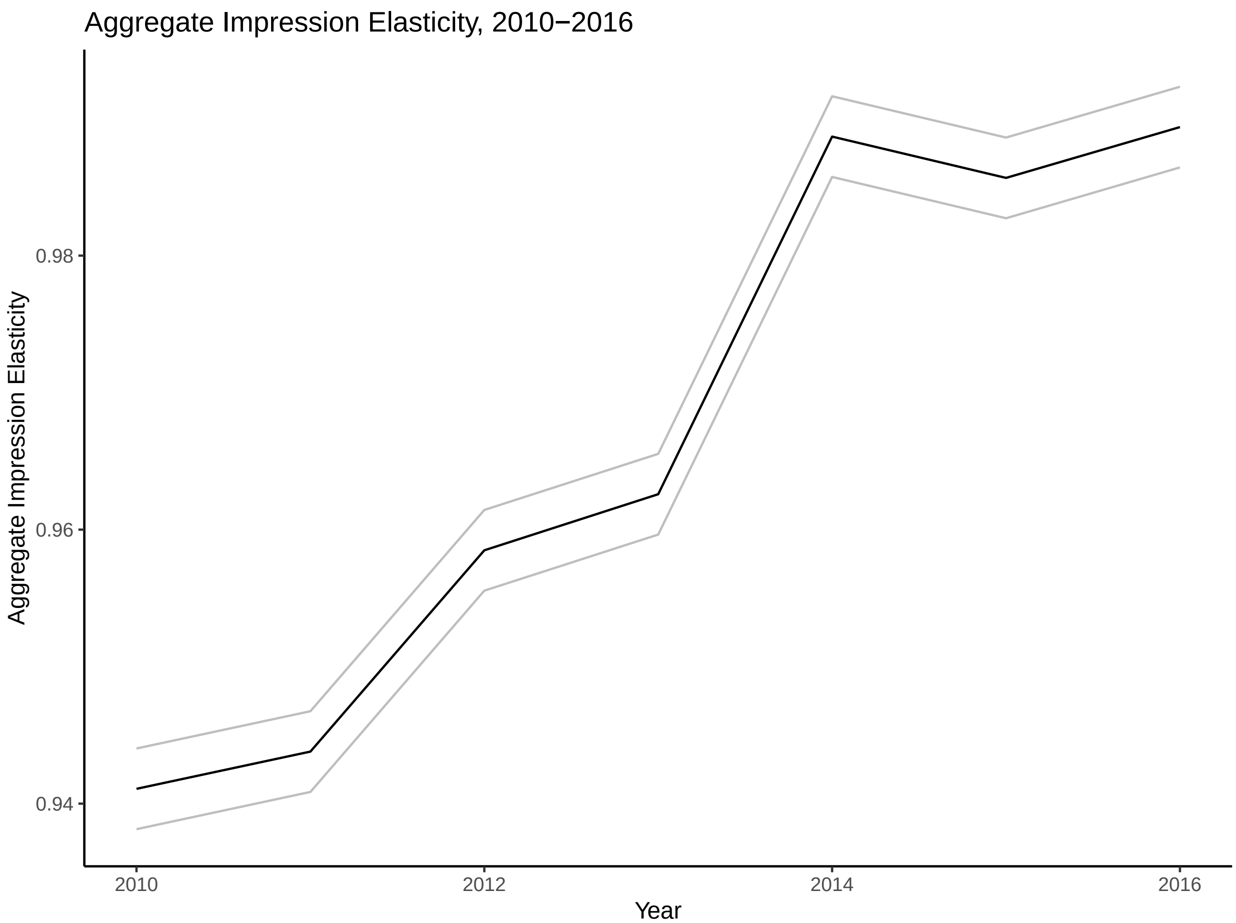


Figure 7: Aggregate impression elasticity, with 95% confidence bands, 2010-2016.

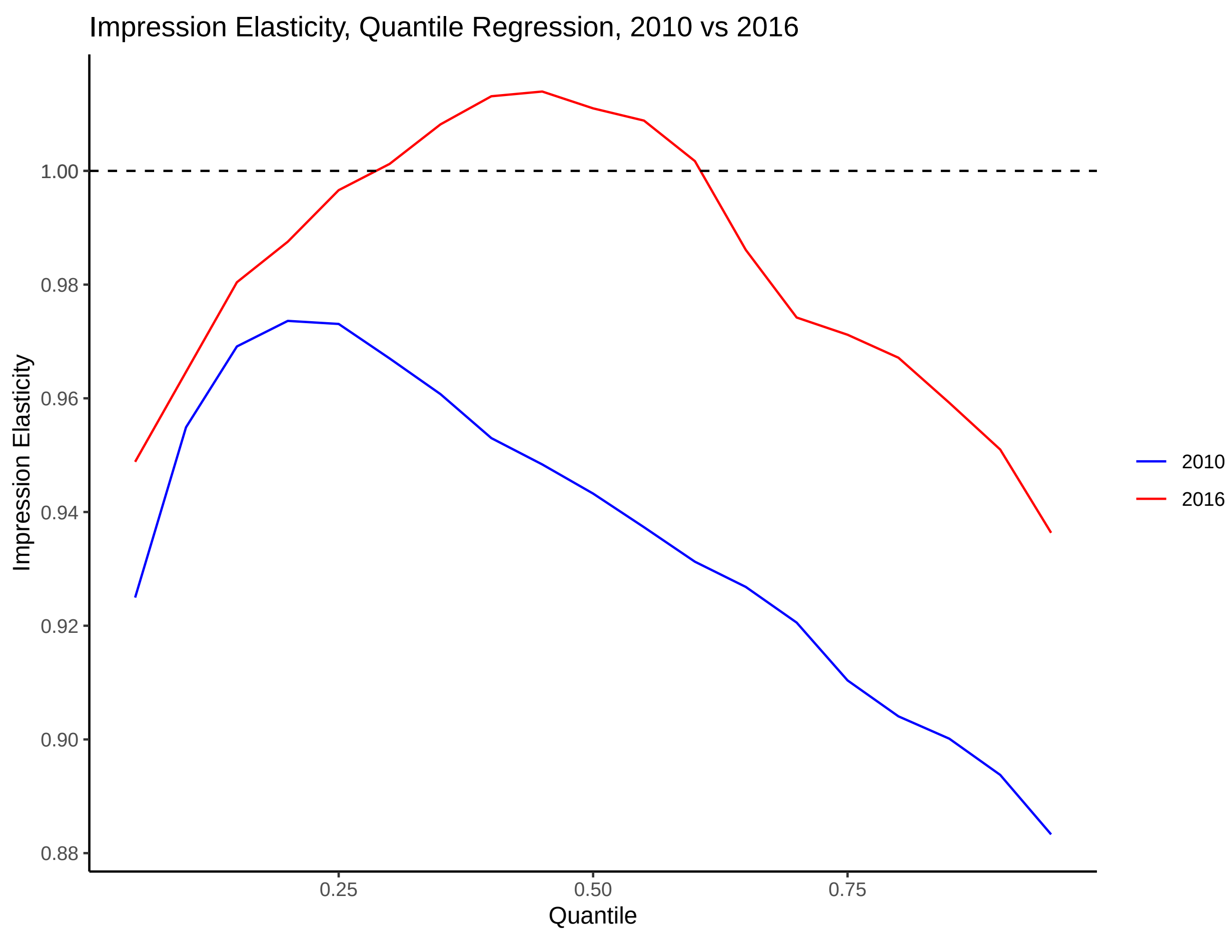


Figure 8: Distribution of impression elasticity in 2010 and 2016, estimated from quantile regressions for 19 quantiles $q = (0.05, 0.1, \cdots, 0.95)$.

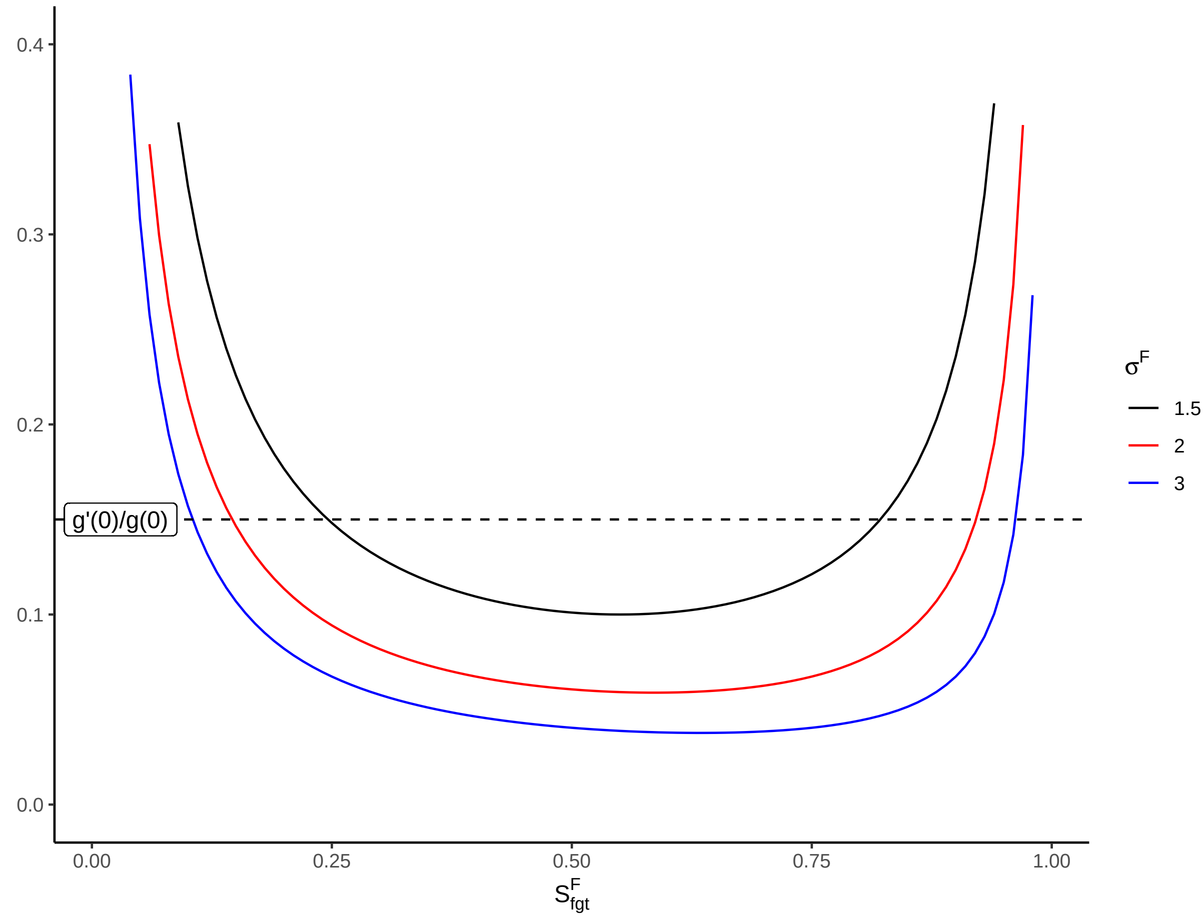


Figure 9: Firm's optimal decision rule for advertising under different values of cross-firm elasticity of substitution, σ^F. The solid curves are the right hand side of (42), when $N^F_{fgt} = 100$ and $E^G_{gt} = 100$. The dotted line is a hypothetical level of $q'(0)/q(0)$. If a firm's sales share is in the region where solid curves are below the dotted line, the firm chooses to advertise in the equilibrium.

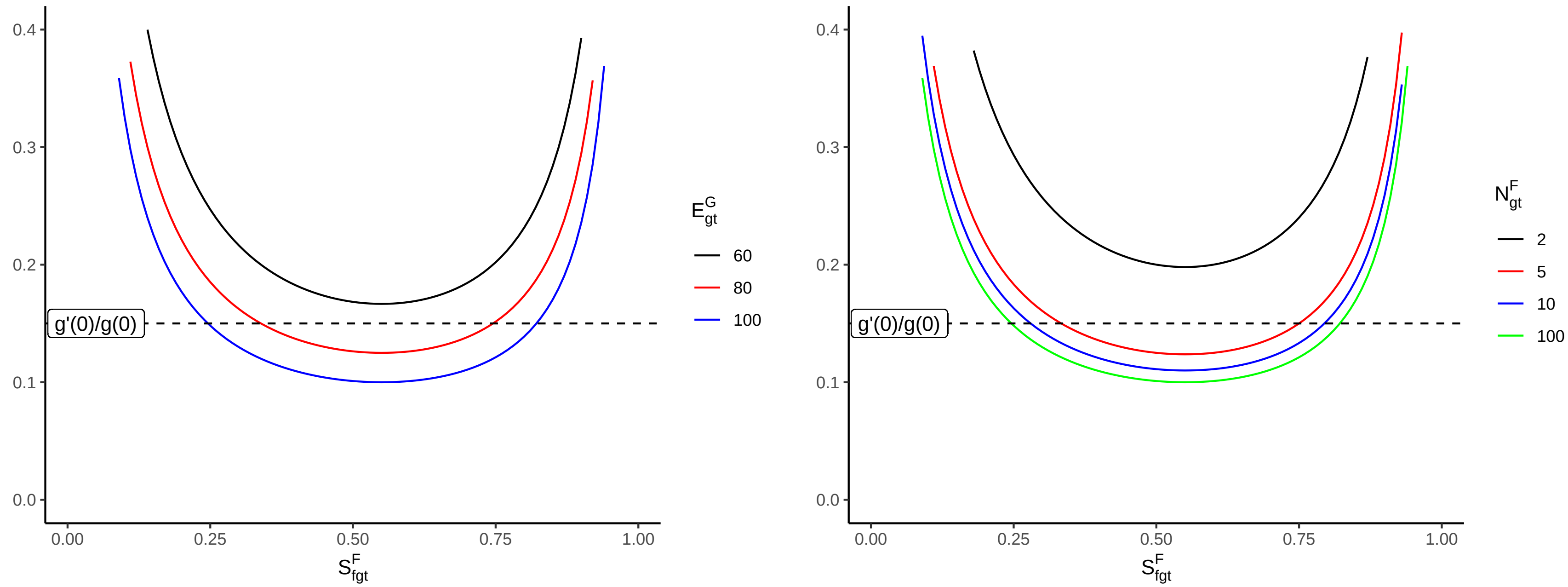


Figure 10: Comparative statics of advertising decision rule under different model parameters, when $\sigma^F = 1.5$. **Left:** $N^F_{gt} = 100$, $E^G_{gt} = \{60, 80, 100\}$, **Right:** $E^G_{gt} = 100$, $N^F_{gt} = \{2, 5, 10, 100\}$. In larger or more competitive product groups, more firms choose to advertise.

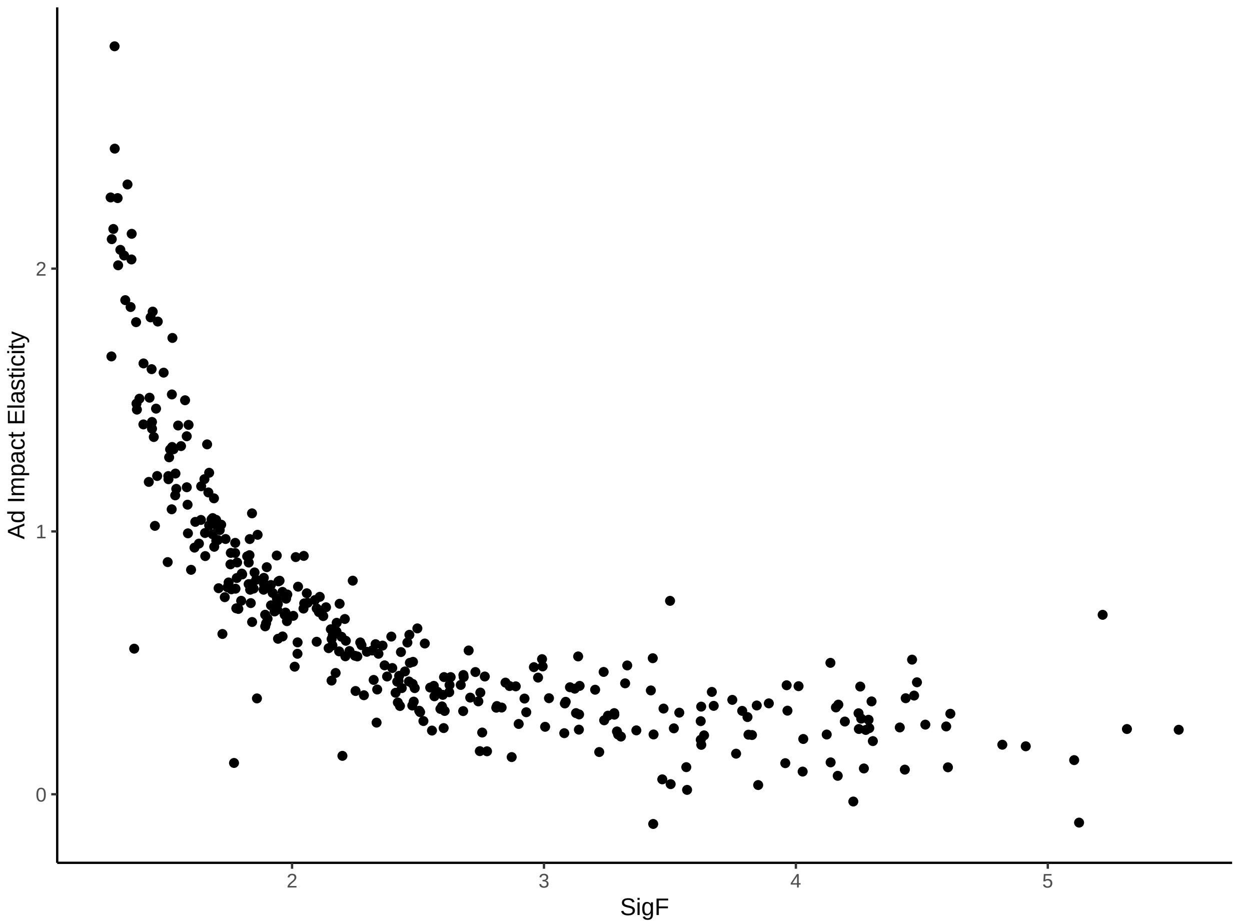


Figure 11: Estimated advertising impact elasticities β_g across product modules, ordered by the elasticity of substitution across firms σ_g^F in each product module. Every dot on the graph represents a unique product module.

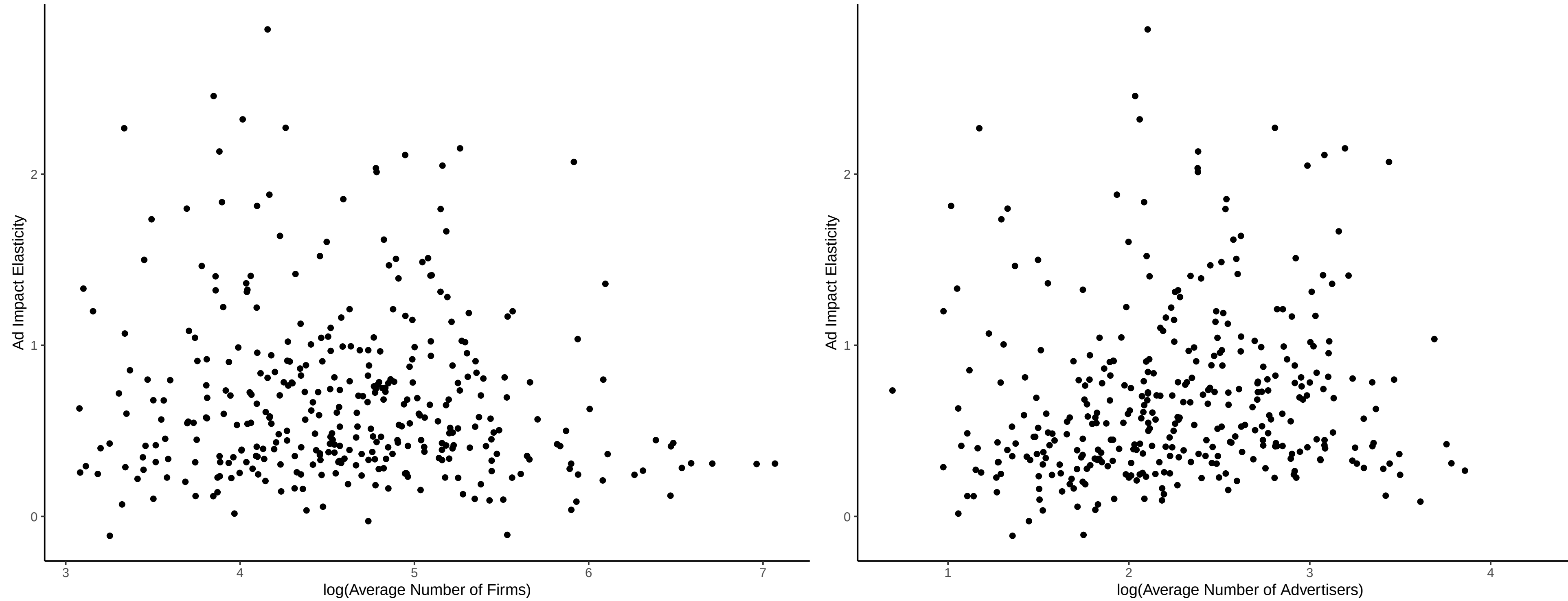


Figure 12: Estimated advertising impact elasticity β_g across product modules, ordered by number of firms (Left) or number of advertisers (Right) in each product module.

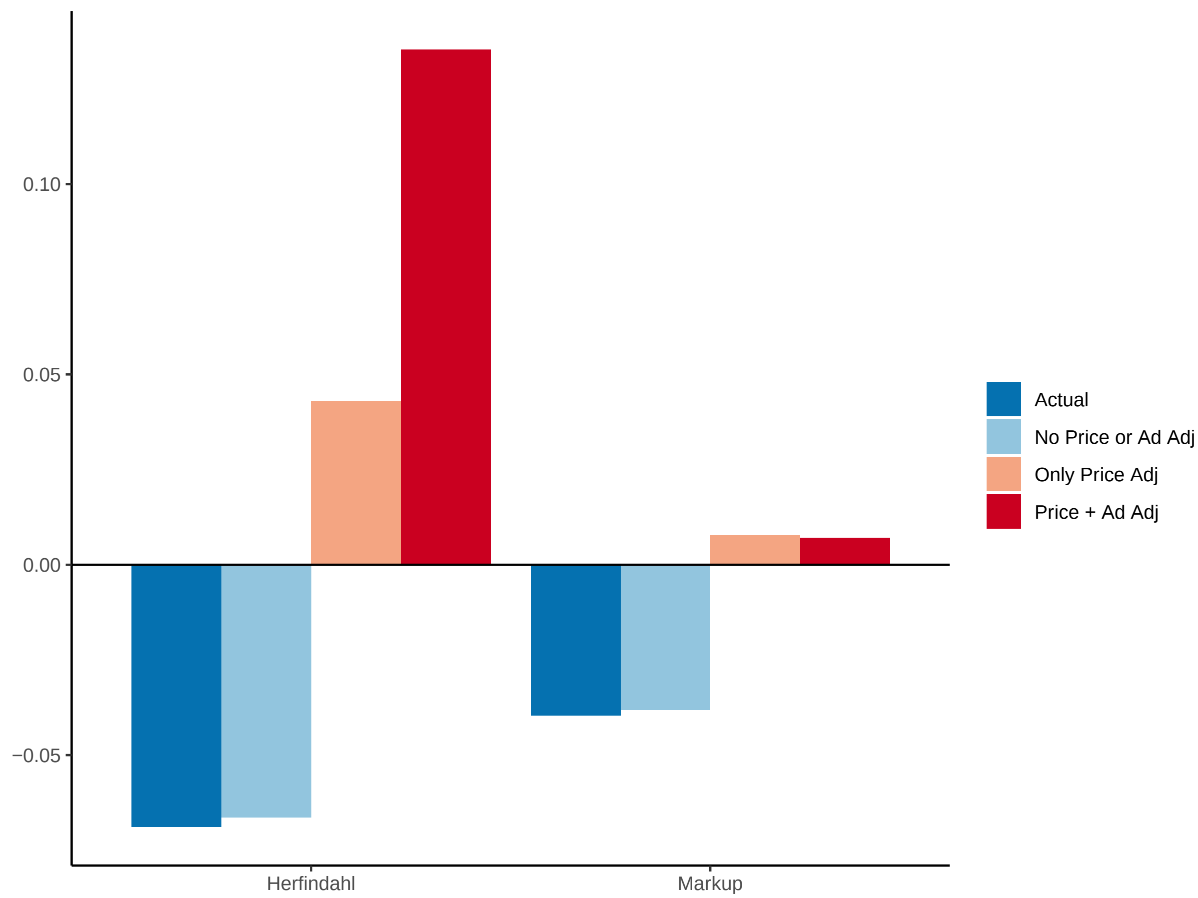


Figure 13: Counterfactual changes in Herfindahl indexes and markup, from 2010 to 2016, if advertising technologies are fixed to the 2010 level. I compare three cases when 1) firms cannot adjust either prices or advertising spending, 2) when they can only adjust prices but not advertising, and 3) when they can freely adjust both.

Table 1: Total Advertising Expenditures in Nielsen Ad Intel, 2010-2015

	Year						
Exp ($M), by Media	2010	2011	2012	2013	2014	2015	**Total**
Network TV	10515.2	10336.5	10952.4	11217.6	11514.9	10618.6	65155.2
Spot TV	1155.7	1229.1	1068.8	1091.6	961.1	988.6	6494.9
Radio	331.2	322.9	300.2	293.4	232.2	243.3	1723.2
Internet	585.1	579.6	446.9	220.5	111.2	50.8	1994.1
Magazine	4918.7	4654.3	4801.7	5165.7	4677.9	4279.2	28497.5
Newspaper	287.0	225.1	226.7	158.0	121.8	90.0	1108.6
Outdoor	270.9	266.8	256.6	264.7	228.6	243.3	1530.9
FSI Coupon	311.0	288.5	280.2	273.0	257.3	192.5	1602.5
Total	18374.8	17902.8	18333.5	18684.5	18105.0	16706.3	108106.9

Table 2: Summary Statistics for Retailer Scanner Dataset, 2010-2015

	2010	2011	2012	2013	2014	2015	Average
Observations (Billions)	13.15	13.65	13.61	13.8	13.9	14.01	13.69
UPCs (Thousands)	803.9	805.8	818	834.7	834.3	844.4	823.5
Product modules	1085	1081	1105	1113	1115	1118	1103
Retail chains	88	88	81	79	76	73	81
Stores	38285	38522	36001	36321	36033	35510	36779
3-digit Zipcodes	877	877	878	877	877	875	877
States	48	48	48	48	48	48	48
Counties	2542	2546	2547	2561	2587	2550	2556
Transaction value ($ Billions)	229.08	241.17	240.26	242.56	246.05	252.19	241.89

Table 3: Summary Statistics for Household Panel Dataset, 2010-2015

	2010	2011	2012	2013	2014	2015
Panelists	60658	62092	60538	61097	61557	61380
Panelists since 2010	60658	47596	40599	35839	31486	28109
States	49	49	49	49	49	49
Counties	2717	2708	2689	2702	2699	2693
5-digit Zipcodes	15946	15998	15815	15849	15972	15893
Total Purchases ($M)	282.4	312.8	302.7	309.1	321,1	313,8
UPCs	726222	739187	763796	788223	791741	794377
Trips	10798312	11269184	10603660	10511397	10579135	10146207
Retailers	767	772	777	773	813	830
Stores	51179	51787	51705	51472	51421	50599

Table 4: Summary Statistics of Merged Data, 2010-2015

	Advertisers			UPCs			Revenue ($Billions)		
Year	Total	Selected Category	Matched	Total	Matched	Advertiser	Total	Matched	Advertiser
2010	337822	5500	1619	803939	530409	122844	229.08	184.26	55.02
2011	474815	7593	2145	805799	561398	132297	241.17	193.40	59.29
2012	596254	9060	2445	817997	583394	137667	240.26	192.34	59.91
2013	709879	10343	2747	834665	595159	143307	242.56	193.11	61.23
2014	810175	11667	3038	834346	616138	149024	246.05	195.17	63.05
2015	901322	13058	3329	844440	650951	163313	252.19	202.19	66.29

Table 5: Baseline Estimation Results for Advertising Elasticity

	Dependent variable:		
	$\log(\text{Sales}_t)$		
	OLS	*Linear Panel*	*GMM (Arellano-Bond)*
	(1)	(2)	(3)
$\log(\text{Ad}_t)$	0.478^{***}	0.081^{***}	0.093^{***}
	(0.008)	(0.005)	(0.0096)
$\log(\text{Sales}_{t-1})$			0.397^{***}
			(0.0507)
Constant	9.732^{***}	14.940^{***}	8.422^{***}
	(0.112)	(0.717)	(0.756)
Observations	3,872	3,872	3,630
R^2	0.465	0.056	-
Fixed Effects	N	Y	-

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Empirical Estimates for Advertising Impact Elasticity β_g

	Dependent variable:		
	$\log(\text{FirmAppeal}_t)$		
	OLS	*Linear Panel*	*Linear Mixed-effects*
	(1)	(2)	(3)
$\log(\text{Ad}_t)$ - $\overline{\log(\text{Ad}_t)}$	0.612***	0.694***	0.697***
	(0.003)	(0.004)	(0.026)
Constant	2.372***		1.720***
	(0.011)		(0.062)
Observations	107,001	107,001	107,001
R^2	0.229	0.259	-
Fixed Effects	N	Y	Y
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7: Summary Statistics of Estimated β_g across Categories

1%	**0.027**	Mean	**0.70**
5%	**0.145**	Number of categories	**356**
25%	**0.340**	Number of advertisers	**2134**
Median	**0.566**	Number of quarters	**32**
75%	**0.903**	Avg categories per firm	**4.33**
95%	**1.751**	Avg firms per category	**137.4**
99%	**2.270**	Avg advertisers per category	**26.0**