Advertising Motives and Firm Life-Cycle Dynamics in a General Equilibrium Model

Miquel Oliver-Vert

University of Nottingham

miquel.oliverivert.nottingham.ac.uk

Latest draft here

October 31, 2024

Abstract

I develop a dynamic general equilibrium model of advertising where consumers are characterised by the subset of goods they are aware of, which evolves over time, and consumers are exposed to advertising while consuming media goods. Firms advertise for three motives: to persuade customers to spend more, to acquire new customers, and, given that consumers' attention is limited, to prevent consumers from learning about competitors. I study how these motives change over the firm life cycle and their aggregate effects, with the informative motive being stronger in younger firms and the persuasive and anticompetitive motives being stronger in mature firms. In the calibrated model, the informative motive is responsible for half of total advertising expenditure. Advertising has a quantitatively significant positive effect on consumption, as consumers enjoy more variety. I also compare the decentralised equilibrium with the planner's allocation. A novel feature of the model is that the planner values media goods even when their entertainment value is negligible, as they serve as a vehicle for product awareness. Finally, I find that advertising should be subsidized, although the gains are small.

1 Introduction

Advertising has long been widely used by firms as a tool both to build their customer base and to increase or retain market share. It can play an informative role, mitigating the information frictions consumers face, enabling them to enjoy more variety and fostering competition. Alternatively, advertising can function persuasively, enhancing consumer preferences for specific goods. This can positively affect consumer utility but may also increase market power by reinforcing product differentiation. An extensive literature has examined whether advertising is informative or persuasive, with supporting evidence for both views.[1] However, research on the macroeconomic implications of advertising remains limited, and a framework accommodating both views is missing. This paper aims to fill this gap.

In addition to these traditional views, consumers' limited attention capacity implies that firms need to compete for the attention of consumers to make their way into their consumption sets. This introduces a novel effect of advertising: advertising by one firm diverts consumers' attention away from competitors. This effect seems particularly relevant in settings like Google search or Amazon advertising, where firms compete to be placed in the top positions within a keyword, as these receive most of the attention.

Since advertising has these three effects and firms may benefit differently from each, some firms may be more motivated to advertise due to one effect than another. Throughout the paper, the terms (i) *informative motive*, (ii) *persuasive motive*, and (iii) *anticompetitive motive* refer to a firm's incentive to advertise in order to (i) inform consumers, (ii) increase spending by current customers, and (iii) reduce consumers' attention to competitors. The term 'anticompetitive' is used to indicate that, under this motive, the firm's advertising aims to avoid competition by hindering competitors' ability to expand their customer base.

What are the aggregate implications of the different advertising motives, and how do these motives evolve over the firm life cycle? This paper develops a novel model that is able to speak about these questions within a general equilibrium framework. Additionally, it is the first to solve a general equilibrium model of informative advertising where asymmetric firms play strategically in a dynamic game, and we believe the methodology developed here will be valuable for subsequent research.

In the model, each industry is composed of a generic good produced under perfect competition by a fringe and an endogenous discrete set of oligopolistic firms each producing a differentiated good. A key feature of the model is that, within an industry, consumers are characterized by the set of goods they are aware of, which I refer to as awareness sets. All consumers know the generic good of each industry, but may not know all of the differenti-

[1] See Bagwell (2007) for a review.

ated goods. The awareness sets evolve stochastically, affected by firms' advertising decisions. Thus, firms face a dynamic problem, as building a customer base takes time, which is motivated by the empirical evidence from Foster et al. (2016).[2] There is one type of advertising that has two effects. First, it increases the probability a consumer becomes aware of the good, and second, it increases the demand shifter for those consumers already aware of the good, inducing them to spend more on the advertised good. Given that the advertising space of each industry is limited, firms internalize that by increasing their advertising expenditure they increase the price of the advertising space in their industry, which reduces the advertising space acquired by competitors, reducing their visibility (the anticompetitive motive). This aligns with Google search advertising, where firms bid on specific keywords in a cost-per-click (CPC) auction.[3] So, heterogeneity in the demand for advertising in a specific keyword translate into heterogeneity in prices.[4]

Additionally, advertising plays a crucial role in the real world by providing revenue for the media goods that consumers tend to enjoy at zero monetary price, spanning traditional outlets like radio and TV as well as digital platforms such as YouTube, Instagram, Facebook, and Google. Although these media goods are barely reflected in GDP (see Greenwood et al., 2024), the time consumers spend on them suggests they have a significant impact on welfare.[5] In the model, consumers choose the time they spend on media based on their entertainment value, and during this time they are exposed to advertising.

The paper studies how the three motives change over the firm life cycle. Intuitively, younger firms have more potential customers to acquire, and so they tend to have a stronger informative motive. In contrast, the anticompetitive motive, which is about retaining market power over the existing customers by reducing the probability they learn about competitors, is stronger for older firms, as they have more to lose due to their larger customer base. The persuasive motive also tends to be stronger in older firms. Intuitively, if advertising persuades existing customers to spend more, the revenue increase will be larger the bigger the customer base.

[2]In particular, Foster et al. (2016) take advantage of data on physical quantities in industries that are plausibly little subject to quality differentiation and find that the fact that older firms are bigger than younger firms cannot be explained by differences in productivity, and then they find support for the hypothesis that firms play an active role, not just a passive effect from aging. Einav et al. (2022), focusing on the retail sector, find that most of the variability in sales is accounted by the number of clients.

[3]Google doesn't charge firms just to be placed at the top, instead, the CPC is the amount the firm will be charged for each click their ad receives. Therefore, Google doesn't necessarily place the highest bidding firm at the top; it also considers the relevance of the ad.

[4]This heterogeneity is shown in Figure 9. The same applies if we look at CPC in Google shopping ads across industries, although these are considerably cheaper, rarely more than one dollar per click.

[5]According to Statista, the average daily time spent on media in the United States in 2023 amounted to 751 minutes.

The model is estimated by simulated method of moments to fit key empirical patterns regarding (i) the evolution of average firm growth by age, which is important to discipline the customer base building process in the model; (ii) the relationship between advertising expenditure and sales, which, given the previous results, is informative of the strength of the motives; and (iii) macroeconomic aggregates, specifically the sales-weighted average markup and its standard deviation, aggregate advertising expenditure as a share of GDP, the fraction of time spent on media, and the labor share. The model does well in matching the targets. In addition, the model also features an inverted-U relationship between advertising and sales as documented in previous literature.

To assess the aggregate effects of the motives and advertising as a whole, the calibrated model is compared to counterfactual economies where some or all of the motives are shut down from the firms' first-order conditions. The counterfactual where all motives are shut down (i.e., there is no advertising) reveals that advertising has a significant positive effect on the aggregate, not only through its role in financing media goods that provide entertainment but also through its role in spreading product awareness. In this counterfactual, the probability consumers learn about goods is lower, as they only learn through an exogenous probability; consequently, firms tend to be smaller. This lowers firms' growth prospects and discourages new firms from entering the market. As a result, consumption decreases, both due to consumers enjoying less variety and because a larger share of consumption comes from a fringe of small, unproductive firms. Overall, shutting down advertising would decrease consumption by 16.68%. The counterfactuals also show that while the persuasive motive increases markups and reduces entry, it has a net positive effect by increasing consumers' taste for the advertised good. The anticompetitive motive is detrimental to output, resulting in higher markups and lower entry. However, it can still have a positive aggregate effect, due to its contribution to the provision of media goods. In fact, a result of the current version of the model is that both the anticompetitive and persuasive motives matter mostly through the entertainment value of media goods, as their effects on consumption are modest. This is despite the fact that these motives significantly influence firms' advertising decisions: shutting down the anticompetitive motive reduces total advertising expenditure by 12%, and additionally shutting down the persuasive motive reduces it by a further 42%. A complementary decomposition exercise based on the firms' first-order condition shows that 8.45% of the (marginal) incentives to advertise are attributable to the anticompetitive motive, 33.74% to the persuasive motive, and 57.81% to the informative motive.

To quantify the inefficiencies, the decentralized equilibrium is compared to the one resulting from solving the social planner problem, while maintaining the consumers' information frictions. A novel feature of the model is that the social planner values media goods not only because they entertain consumers, as in existing literature, but also because, through the advertising in media, consumers get information that allows them to improve consumption. In other words, media serves as a vehicle for product awareness. Unsurprisingly, as the entertainment value of media goods increases, the social planner reallocates more labor from the production sector to the media sector. After a certain point, consumption under the planner's allocation becomes lower than under the decentralized one. More interestingly, when the entertainment value of media is negligible, the optimal quantity of media is lower than in the decentralized equilibrium, suggesting excessive advertising expenditure. However, this conclusion would be inaccurate as we must also consider how the advertising space is allocated among firms. In other words, the 'overprovision' of media, through its effect on learning, may help mitigating the inefficiencies arising from the misallocation in the advertising space. In this direction, the exercise examining the optimal uniform tax on advertising reveals that advertising should be subsidized.[6] Finally, given that the informative motive declines with firm age, while the persuasive and anticompetitive motives increase, a natural question to ask is what the welfare gains from allowing the advertising tax to be age-dependent would be. However, in the current version, the gains from such a policy are negligible.

The paper is organized as follows: Section 2 introduces the model and characterize the equilibrium. Section 3 estimates the model, studies the evolution of the motives, their contribution to total advertising expenditure, and their aggregate effect. Section 4 discusses the inefficiencies of the comed, compares the informationally-constrained social optimal equilibrium to the decentralized one, and examines the gains from taxing advertising. Finally, section 5 concludes.

Related literature. This paper relates to the literature that studies the implications of customer capital for firm, industry, and macroeconomic dynamics (e.g. Dinlersoz and Yorukoglu (2012), Gourio and Rudanko (2014), Molinari and Turino (2018), Argente et al. (2023), Einav et al. (2022), Ignaszak and Sedlacek (2023), Greenwood et al. (2024)). In these models, firms grow via increasing their idiosyncratic demand (customer capital). Together with Cavenaile et al. (2024b), we contribute to this literature by showing that it is not just the quantity of customers that matters, but also the degree of information the customers have about alternative goods. Relative to Cavenaile et al. (2024b), this paper allows for strategic advertising decisions as well as the interaction between firms of different sizes and ages. In Cavenaile et al. (2024b), advertising also serves to expand product awareness, but the

[6] In this exercise, I compare the stationary equilibria resulting from the different tax levels, without considering the transition.

advertising choices are coordinated at the industry level, made once and for all at industry inception, and firms are assumed to be symmetric. Their focus is on how the improvements in targeted advertising may lead to increased market power through market segmentation.

This paper is also related to Greenwood et al. (2024), who present a static model to study the inefficiencies from advertising. Here, the contribution is to add to the analysis the inefficiencies arising from markups typical of oligopoly frameworks, as well as identifying three novel sources of inefficiency from advertising. The first two are reminiscent of growth models, namely: (i) lack of full appropriability, as the producers cannot extract the entire surplus; and (ii) business-stealing, as firms don't consider the losses from the reduction of consumption from competitors. The third source of inefficiency arises from the anticompetitive motive particular to this model, as firms try to avoid suffering from the business-stealing effect. Note that the sources of inefficiency push in different directions, so it is not clear whether there is too much or too little advertising, and requires a quantitative answer. Finally, there is inefficient entry, again due to lack of full appropriability and business-stealing.

For the persuasive aspect of advertising, this paper builds on the literature that adopts the persuasive view, e.g. Cavenaile et al. (2024a), Rachel (2024), Molinari and Turino (2018). These papers model advertising as a static demand shifter. A novel contribution of the current paper is the combination of the persuasive and informative views of advertising within a single framework. In particular, this paper relates to Cavenaile et al. (2024a), as they also develop an oligopolistic model with endogenous market structure. In their paper, they study the interaction between R&D and advertising. As in my model, advertising in Rachel (2024) and Greenwood et al. (2024) also finances the provision of media goods that improve utility.

Finally, this paper uses the concept of consideration sets introduced by Manzini and Mariotti (2014) that are widely used in other fields.[7] In macroeconomics, this concept (using the term *awareness set*) is introduced by Cavenaile et al. (2024b). Relative to them, this paper explores how the evolution of awareness sets both influences and is influenced by firms' advertising decisions. The presence of awareness sets complicates the firm problem, as firms need to keep track of the distribution of consumers across these sets. In their model, they abstract from this by assuming the evolution of the awareness sets is determined at industry inception, so the only state variable is industry age.

[7] Manzini and Mariotti (2014) model choice as a two-stage process. In the first stage, some of the available alternatives are selected into a *consideration set*, with a probability that is linked to attention. In the second stage, the agent maximizes utility restricted to the consideration set.

2 Model

2.1 Environment

Market structure and the production sector. There is a continuum of mass 1 of industries indexed by i. In each industry, there is a generic good and a discrete set $\mathcal{J}_{i,t}$ of firms, indexed by j, each one producing a single differentiated good with the production function $y_{j,i,t} = N_{j,i,t}$, where $N_{j,i,t}$ is the labor employed by firm j. The generic good is produced under perfect competition by many small firms with the production function $y_{0,i,t} = A_0 N_{0,i,t}$, where $N_{0,i,t}$ is the total labor employed by these small firms in industry i at period t.

Advertising and the media sector. There is a media sector populated by media firms that employ labor to produce media goods, which are supplied to consumers at zero monetary price, and generate revenue by selling advertising space to production firms. There is free entry. Each media firm produces a differentiated variety of media good of equal quality, so consumers will allocate the time they spend on media equally among the different media goods. The aggregate quality of media is given by

$$Q = AN_m^{\frac{1}{2}} \tag{1}$$

where N_m is total labor employed in media. In order to rule out an equilibrium with no advertising expenditure and no media produced, I assume the government employs $\bar{N}_m$ units of labor in media, which is financed by a lump-sump tax to consumers.
Within the media sector, each industry of the production sector has α units of advertising space.[8] The process whereby firms acquire advertising space follows a kind of auction, where media firms post a price per unit of ad space in industry i, $p_{a,i,t}$, which is the minimum bid accepted, and supply at most α units of ad space. Letting $e_{j,i,t}$ be the advertising expenditure of firm j in industry i, then the final price per unit of ad space in industry i will be equal to $\max\left\{p_{a,i,t}, \sum_{j\in\mathcal{J}_{i,t}} e_{j,i,t}/\alpha\right\}$. Therefore, the advertising space acquired by firm j, $\alpha_{j,i,t}$, will be:

$$\alpha_{j,i,t} = \min\left\{\frac{e_{j,i,t}}{p_{a,i,t}}, e_{j,i,t}\frac{\alpha}{\sum\limits_{k\in\mathcal{J}_{i,t}} e_{k,i,t}}\right\} \tag{2}$$

[8] This is a reasonable assumption for search advertising: there is one top search position for a specific keyword, so higher demand only leads to higher price, as suggested in Figure 9. More generally, you could think of this α as some measure of attention.

Entry and Exit of firms. A firm is hit by a death shock with probability κ, independent of whether other firms are affected (so, the probability n firms exit is κ^n).[9] Regarding entry, there is a measure one of entrepreneurs that employs $N_{e,i,t}$ units of labor to create a new differentiated good in industry i with probability $z_{e,i,t} = \phi_s N_{e,i,t}^{\frac{1}{2}}$. Upon successfully creating a new good (and, for computational purposes, provided the number of firms in the industry is below $\bar{J}$), a new firm enters the market, and initially no consumer is aware of the new firm. Entry and exit occur simultaneously right at the start of $t+1$.

Consumers. There is a unit mass of individuals indexed by ℓ who maximize lifetime utility, where the instantaneous utility is a function of her consumption (C_ℓ) and entertainment (L_ℓ) goods. Individuals die with an exogenous probability δ, in which case they are replaced with an offspring who inherits the assets $a_{\ell,t}$, and individuals discount the offspring's utility with the same discount rate; thus, we can write utility as if they were infinitely lived:[10]

$$U_\ell = \sum_{t=0}^{\infty} \beta^t \; [\mathbb{E} \ln C_{\ell t} + L_{\ell t}] \tag{3}$$

Each individual supplies inelastically one unit of labor and chooses how much time to allocate to media goods, $T_{\ell,t}$, in order to maximise her entertainment good $L_{\ell,t}$, which is defined as follows:[11]

$$L_{\ell t} = \upsilon \left(Q_t T_{\ell,t} - \frac{T_{\ell,t}^2}{2} \right) \tag{4}$$

where Q_t is an output of the media sector production function. Anticipating that all individuals choose the same $T_{\ell,t}$, in what follows I drop the subindex ℓ from T_t and L_t. Individual ℓ gets her C_ℓ following a Cobb-Douglas aggregator of her consumption over the continuum of industries of mass 1

$$\ln C_{\ell,t} = \int_0^1 \ln C_{\ell,i,t} di \tag{5}$$

where the industry i consumption good of individual ℓ is a CES aggregator of her consumption on the generic good and each of the differentiated goods she is aware of:

[9] I plan to do an extension where this probability is decreasing in firm size. This is a more realistic assumption, and would likely imply a stronger anticompetitive motive to advertise, as preventing competitors from expanding effectively increases their probability of exiting.

[10] Note that there is no uncertainty on $C_{\ell,t}$. There is uncertainty at the industry level due to the stochastic evolution of the awareness sets (see next section), but the law of large numbers over the continuum of industries removes the uncertainty at the aggregate level.

[11] Note that $T_{\ell,t}$ is not restricted to be below 1; this is consistent with the way media time is measured in the data where multitasking is counted separately, see Appendix 6.1.

$$C_{\ell,i,t} = \left(c_{\ell,0,i,t}^{\frac{\sigma-1}{\sigma}} + \sum_{j \in \mathcal{I}_{\ell,i,t}} \omega_{j,i,t} c_{\ell,j,i,t}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1 \tag{6}$$

where $c_{\ell,j,i,t}$ is the quantity of good j consumed by ℓ at t; $\mathcal{I}_{\ell,i,t}$ will be referred to as the awareness set of individual ℓ in industry i at period t, as it is the subset of the differentiated goods $\mathcal{J}_{i,t}$ the individual is aware of at period t (in the next section, I describe the evolution of this object); and $\omega_{j,i,t}$ is a demand shifter that depends on the exposure of individuals to the ad of good j. In particular:

$$\omega_{j,i,t} = 1 + \nu_s(\alpha_{j,i,t} T_t)^{\nu_c}, \quad \nu_c \in (0,1) \tag{7}$$

$\omega_{j,i,t} = 1 + \nu_s(\alpha_{j,i,t} T_t)^{\nu_c}$, $\nu_c \in (0,1)$. Note that the more time consumers spend on media, the larger the effect of advertising on the demand shifter.

Individual ℓ's budget constraint writes:

$$w_t N_{\ell,t} + r_t a_{\ell,t} = \int_0^1 \sum_{j \in \mathcal{I}_{\ell,i,t} \cup \{0\}} c_{\ell,j,i,t} p_{j,i,t} di + a_{\ell,t+1} - a_{\ell,t} + \tau_t \tag{8}$$

where w_t is the wage, $a_{\ell t}$ is the asset holding of individual ℓ at period t, r_t is the return on each unit of asset in period t, and τ_t is the lump-sum tax the government uses to employ $\bar{N}_m$ units of labor in the media sector. All individuals start with the same level of assets a_0.

Product learning and the evolution of the awareness sets. I assume that the probability an individual gets aware of a product thanks to advertising is an increasing and concave function of the exposure to the ad of that good. In particular, assume an individual will get aware of product j in industry i with probability[12]

$$\rho_{j,i,t} = \hat{\rho} + \psi_s(\alpha_{j,i,t} T_t)^{\psi_c}, \quad \psi_c \in (0,1) \tag{9}$$

Although I focus on advertising as an active way through which firms can increase their customer base, consumers can get to know a firm in other ways (word-of-mouth, seeing the product in a shop...), and these are captured by $\hat{\rho}$. The inclusion of T_t is to capture the idea that the more time consumers spend on media, the more they are exposed to ads, and so the more effective advertising is, just like in the demand shifter $\omega_{j,i,t}$.

[12] In a version of the model, I have congestion, $\alpha_i^{-\zeta}$, $\zeta \in [0, \psi_c]$, which allows to play with the intensity of the anticompetitive motive.

The events of learning goods are assumed to be independent; that is, the probability of learning all the goods in $\mathcal{I} \subseteq \mathcal{J}_i$ is $\prod_{j\in\mathcal{I}} \rho_{j,i}$. And, given that the complements of independent events are also independent, then the probability of not learning any of the goods in $\mathcal{I} \subseteq \mathcal{J}_i$ is $\prod_{j\in\mathcal{I}}(1-\rho_{j,i})$.

In addition, when a consumer dies, they are replaced by a newborn individual who starts knowing only the generic good of each industry (i.e. $\mathcal{I}_{\ell,i} = \emptyset$ for all i). This is completely equivalent to say that individuals forget all the differentiated goods they know with an exogenous probability δ. This assumption is not crucial for the results, and its only implication is that, even if a firm lived forever, there would always be some consumers that are not aware of it.

Then, we have all the information needed to find the probability of moving between any pair of awareness sets. Let $\Theta_{(\mathcal{I}\to\mathcal{I}')}$ be the probability of moving from $\mathcal{I}$ to $\mathcal{I}'$. Given that, conditional on not dying, the awareness set can only expand, then if $\mathcal{I}'$ doesn't contain $\mathcal{I}$, the transition is only possible (i.e. $\Theta_{(\mathcal{I}\to\mathcal{I}')} > 0$) if $\mathcal{I}' = \emptyset$, which happens with the probability of dying δ. Conversely, if $\mathcal{I}'$ contains $\mathcal{I}$, then the probability this transition takes place is the probability an individual doesn't die, $(1-\delta)$, times the probability of learning all the goods that are in $\mathcal{I}'$ but not in $\mathcal{I}$, $\prod_{j\in\mathcal{I}'\setminus\mathcal{I}} \rho_{j,i}$, times the probability of not learning any of the goods that are not in $\mathcal{I}'$, $\prod_{j\notin\mathcal{I}'}(1-\rho_{j,i})$. Formally:

$$\Theta_{(\mathcal{I}\to\mathcal{I}')} = \begin{cases} 0, & \text{if } \mathcal{I} \nsubseteq \mathcal{I}' \neq \emptyset \\ \delta, & \text{if } \mathcal{I} \nsubseteq \mathcal{I}' = \emptyset \\ (1-\delta) \prod_{j\in\mathcal{I}'\setminus\mathcal{I}} \rho_{j,i} \cdot \prod_{j\notin\mathcal{I}'} (1-\rho_{j,i}), & \text{if } \mathcal{I} \subseteq \mathcal{I}' \neq \emptyset \\ (1-\delta) \prod_{j\in\mathcal{J}_{i,t}} (1-\rho_{j,i}) + \delta, & \text{if } \mathcal{I} = \mathcal{I}' = \emptyset \end{cases} \tag{10}$$

2.2 Equilibrium

In this section, I characterize the pure strategy Markov perfect stationary equilibrium, that is such that the time spent in media T_t and the relative wage $\hat{w}_t = \frac{w_t}{E_t}$ are constant.

2.2.1 Consumption.

On the one hand, logarithmic preferences on $C_{\ell,t}$, together with $a_{\ell,0} = a_0$, imply that all consumers choose the same expenditure at all t: $E_{\ell,t} = E_t$. On the other hand, CES preferences over the varieties within an industry implies that consumer's spending in an

industry is independent of her industry price index: $E_{\ell,i,t} = E_t$. Therefore, the awareness sets $\mathcal{I}_{\ell,i,t}$ only affect the allocation of the expenditure within each industry. That is, in order to characterize consumer ℓ's consumption choices in industry i, we only need to know her awareness set in i, $\mathcal{I}_{\ell,i,t}$. In other words, within industry i, there are as many types of consumers as subsets $\mathcal{I} \subseteq \mathcal{J}_{i,t}$. So, the set of consumer types in industry i is identified by the power set of $\mathcal{J}_{i,t}$, $\mathcal{P}(\mathcal{J}_{i,t})$, and, within an industry, I'll use subindex $\mathcal{I}$ to denote the choice of an individual with awareness set $\mathcal{I}$.
The optimal choices satisfy: [13]

$$c_{\mathcal{I},j,i,t} = E_t P_{\mathcal{I},i,t}^{\sigma-1} p_{j,i,t}^{-\sigma} \omega_{j,i,t}^{\sigma}, \quad j \in \mathcal{I} \tag{11}$$

$$\frac{E_{t+1}}{E_t} = \beta(1 + r_{t+1}) \tag{12}$$

$$T_t = Q_t \tag{13}$$

where $P_{\mathcal{I},i,t} = \left(p_{0,i,t}^{1-\sigma} + \sum_{j \in \mathcal{I}} \omega_{j,i,t}^{\sigma} p_{j,i,t}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$.[14] Note that consumers consume a positive amount of all the goods they are aware of, and the particular quantity consumed follows equation (11). Equation (12) is the Euler equation, and (13) states that the aggregate quality of media determines the time spent on media. From $E_t = P_{\mathcal{I},i,t} C_{\mathcal{I},i,t}$ we see one of the channels through which (the informative) advertising will increase welfare: advertising will increase the amount of products the consumer is aware of, which reduces the price $P_{\mathcal{I},i,t}$ of her industry composite good.This is a standard love for variety effect.

2.2.2 The industry state and its evolution

Given that firms have the same production technology, all the heterogeneity comes from the consumer side. The relevant state of the industry is characterised by the triple $(\mathcal{J}_{i,t}, \mathcal{P}(\mathcal{J}_{i,t}), \vec{M}_{i,t})$, where $\vec{M}_{i,t} = (M_{i,t}(\mathcal{I}))_{\mathcal{I} \in \mathcal{P}(\mathcal{J}_{i,t})}$ is the mass of consumers in each awareness set (that is, how consumers are distributed over the awareness sets). Note that since there is a mapping from $\mathcal{J}_{i,t}$ to $\mathcal{P}(\mathcal{J}_{i,t})$, I may write the state simply as $(\mathcal{J}_{i,t}, \vec{M}_{i,t})$. There are two processes that shape the evolution of the industry state.
On the one hand, the industry state changes as consumers' awareness sets evolve due to learning and death, which, by law of large numbers, is a deterministic process at the in-

[13] Together with the No-Ponzi condition $\lim_{\tau \to \infty} \frac{a_{t+\tau}}{\prod_{s=0}^{\tau}(1+r_{t+s})} = 0$.

[14] Note that consumers may not only have different industry price index, $P_{\ell,i,t}$, but also a different aggregate price index $P_{\ell,t} = exp\left(\int_0^1 \ln P_{\ell,i,t} di\right)$. In particular, as explained in section 2.2.8, individuals with the same age have the same aggregate price index, and the numeraire of the economy is the geometric mean of $P_{\ell,t}$.

dustry level.[15] Calling Θ_t the transition matrix, where the element in row r and column s indicates the probability of going from subset $\mathcal{I}_r$ to $\mathcal{I}_s$ at time t (i.e. $\Theta_{t,(\mathcal{I}_r \to \mathcal{I}_s)}$), and calling $\vec{M}_{i,t}$ the $2^{\#\mathcal{J}_{i,t}}$-dimensional row vector (where $\#\mathcal{J}_{i,t}$ is the cardinal of $\mathcal{J}_{i,t}$) containing the masses of consumers in each awareness set at time t; then, by the law of large numbers, the distribution of consumers in $t+1$ in the absence of entry and exit of goods, which I denote by $\vec{\hat{M}}_{i,t+1}$, would be:

$$\vec{\hat{M}}_{i,t+1} = \vec{M}_{i,t}\Theta_t \tag{14}$$

On the other hand, the industry state changes stochastically due to entry and exit of firms. If the realisation of exit and entry changes the set of firms in industry i from $\mathcal{J}$ to $\mathcal{J}'$, then the next period industry state is obtained using the application $(\mathcal{J}, \vec{\hat{M}}, \mathcal{J}') \longmapsto (\mathcal{J}', \vec{M}')$ defined as follows.

$$\text{For } \mathcal{I}' \in \mathcal{P}(\mathcal{J}'), \qquad M'(\mathcal{I}') = \begin{cases} \sum\limits_{\{\mathcal{I} \in \mathcal{P}(\mathcal{J}) : \mathcal{I} \cap \mathcal{J}' = \mathcal{I}'\}} \hat{M}(\mathcal{I}) & \text{, if } \mathcal{I}' \subseteq \mathcal{J} \\ 0 & \text{, if } \mathcal{I}' \nsubseteq \mathcal{J} \end{cases} \tag{15}$$

where the first case says that two consumers become identical in industry i if all the firms in which they differed exit, whereas the second case says that there are no consumers who are aware of a newborn firm. The last piece of infornation needed to compute expected values is the probabilities that the set of differentiated goods moves from $\mathcal{J}$ to $\mathcal{J}' \subseteq \mathcal{J} \cup \{e\}$, where e denotes an entrant. These probabilities are given by:

$$\text{For } \mathcal{J}' \in \mathcal{P}(\mathcal{J} \cup e)), \qquad Prob\{\mathcal{J} \to \mathcal{J}'\} = \begin{cases} (1 - z_{e,i,t}) \prod\limits_{j \in \mathcal{J} \cap \mathcal{J}'} (1-\kappa) \prod\limits_{j \in \mathcal{J} \setminus \mathcal{J}'} \kappa & \text{, if } e \notin \mathcal{J}' \\ z_{e,i,t} \prod\limits_{j \in \mathcal{J} \cap \mathcal{J}'} (1-\kappa) \prod\limits_{j \in \mathcal{J} \setminus \mathcal{J}'} \kappa & \text{, if } e \in \mathcal{J}' \end{cases} \tag{16}$$

where $z_{e,i,t}$ is the probability of an entrant, $\prod\limits_{j \in \mathcal{J} \cap \mathcal{J}'} (1-\kappa)$ is the probability that all the firms in $\mathcal{J} \cap \mathcal{J}'$ survive, and $\prod\limits_{j \in \mathcal{J} \setminus \mathcal{J}'} \kappa$ is the probability that all the firms in $\mathcal{J} \setminus \mathcal{J}'$ exit.

In the Appendix 6.5 I show that assuming that individuals don't die (i.e. $\delta = 0$) allows a simpler sufficient industry state given by the vector of customer bases. That is, instead of requiring the mass of consumers in each awareness set, we would only need to know the mass of consumers aware of each good.

[15] In case there were mixed strategies (although this is not the case in the equilibrium studied) this process would be stochastic.

2.2.3 Production firms problem

Given the large number of small firms producing a homogeneous product, the price of the generic good is equal to its marginal cost, $p_{0,i,t} = \frac{w_t}{A_0}$. The differentiated firms compete in prices a la Bertrand and in advertising expenditures for the attention of consumers. Both decisions are made simultaneously.

Profits. Using the production function $y_{j,i,t} = N_{j,i,t}$, we can express profits decomposed as

$$\pi_{j,i,t} = \underbrace{M_{j,i,t}}_{\substack{\text{Customer}\\ \text{Base}}} \cdot \underbrace{(1 - \mathcal{M}_{j,i,t}^{-1}) \underbrace{\sum_{\mathcal{I} \in \mathcal{P}_j(\mathcal{J}_{i,t})} \frac{M_{i,t}(\mathcal{I})}{M_{j,i,t}} s_{\mathcal{I},j,i,t} E_t}_{\substack{\text{Average spending}\\ \text{by customers}}}}_{\text{Average rents from customers}} \tag{17}$$

where $\mathcal{M}_{j,i,t} = \frac{p_{j,i,t}}{w_t}$ is the markup of firm j, $s_{\mathcal{I},j,i,t} = \frac{p_{j,i,t} c_{\mathcal{I},j,i,t}}{E_t}$ is type $\mathcal{I}$ individual's share of expenditure in good j, $\mathcal{P}_j(\mathcal{J}_{i,t}) = \{\mathcal{I} \in \mathcal{P}(\mathcal{J}_{i,t}) \mid j \in \mathcal{I}\}$ is the family of awareness sets containing good j, and $M_{j,i,t} = \sum_{\mathcal{I} \in \mathcal{P}_j(\mathcal{J}_{i,t})} M_{i,t}(\mathcal{I})$ is the customer base of firm j.

This expression offers a first intuition of the motives driving firms to advertise. First, they want to advertise to increase their customer base. I refer to this as the informative motive. Second, as shown in the Appendix 6.6, all else equal, firms prefer to have customers that know as fewer competitors as possible. Intuitively, the fewer alternative goods they know, the more they will spend in j (i.e. higher $s_{\mathcal{I},j,i,t}$) and the lower their demand elasticity (so, the firm is able to extract more rents by rising the markup). So, given that by increasing the advertising space they occupy, firms reduce the attention of consumers to the competitors' goods and so the probability they will add them to their awareness sets; then, firms may have the incentive to do advertising for the mere purpose of reducing the mass of customers who learn about competitors. I refer to this as the anticompetitive motive, as under this motive the firm is doing advertising to avoid competition by precluding competitors to expand their customer base. Finally, given that the demand shifter $\omega_{j,i,t}$ increases with the the advertising space, firms want to do advertising to persuade current consumers to buy more. This is the persuasive motive.

Price setting. I focus on pure strategy Markov perfect equilibrium where policy functions only depend on the current industry state $(\mathcal{J}_{i,t}, \vec{M}_{i,t})$. Given that the price has no direct effect on the evolution of the industry state and that advertising and price choices are made simultaneously, then the price setting problem is static. The optimal markup $\mathcal{M}_{j,i,t}$ is such that profits (17) are maximised, given its own demand-shifter $\omega_{j,i,t}$, the markups and demand-shifters of the competitors $\{\mathcal{M}_{k,i,t}, \omega_{k,i,t}\}_{k\in\mathcal{J}_{i,t},k\neq j}$, and the distribution of consumers over the awareness sets $\vec{M}_{i,t} = (M_{i,t}(\mathcal{I}))_{\mathcal{I}\in\mathcal{P}(\mathcal{J}_{i,t})}$, and taking into account that individuals' spending shares are given by

$$s_{\mathcal{I},j,i,t} = \left[(A_0\mathcal{M}_{j,i,t})^{\sigma-1}\omega_{j,i,t}^{-\sigma} + \sum_{k\in\mathcal{I}}\left(\frac{\omega_{k,i,t}}{\omega_{j,i,t}}\right)^{\sigma}\left(\frac{\mathcal{M}_{j,i,t}}{\mathcal{M}_{k,i,t}}\right)^{\sigma-1}\right]^{-1} \tag{18}$$

The equilibrium markups are given by:

$$\mathcal{M}_{j,i,t} = \frac{\frac{\sigma}{\sigma-1} - \bar{s}_{j,i,t}}{1-\bar{s}_{j,i,t}} \text{ , with } \bar{s}_{j,i,t} = \sum_{\mathcal{I}\in\mathcal{P}_j(\mathcal{J}_{i,t})}\frac{M_{i,t}(\mathcal{I})p_{j,i,t}c_{\mathcal{I},j,i,t}}{p_{j,i,t}y_{j,i,t}}s_{\mathcal{I},j,i,t} \tag{19}$$

where $\bar{s}_{j,i,t}$ is the sales-weighted average of firm j customers' share of expenditure in industry i allocated to good j.

Note that in a standard oligopoly model with Bertrand competition, the optimal markup is given be the expression in (19) but with the market share $s_{j,i,t}$ instead of $\bar{s}_{j,i,t}$. So, while the optimal markup in a standard oligopoly model with Bertrand is increasing with size, this is not necessarily the case here. Here, the markup depends on the composition of the customers, not in the size: a smaller firm can have a higher markup if a larger fraction of its customers spend a larger share of expenditure on it. However, the model will still predict that, within an industry, larger firms have higher markups. The intuition is as follows: a firm that entered earlier had more time to accumulate customers (so older firms will be larger); but also, since as time passes consumers get aware of more goods and advertising is undirected, then a firm that enters later will get consumers that, on average, know more goods (and we have seen that customers with more alternative goods spend a smaller share). So, within an industry, larger firms will have customers that on average spend a larger share of expenditure, and thus they set higher markups.

Advertising choice. Each firm chooses dynamically its advertising expenditure $e_{j,i,t}$, taking into account (i) the advertising expenditure choices of its competitors $\{e_{k,i,t}\}_{k\in\mathcal{J}_{i,t},k\neq j}$; (ii) markups $\{\mathcal{M}_{k,i,t}\}_{k\in\mathcal{J}_{i,t}}$; (iii) the time consumers spend on media, T_t; (iv) the law of motion of the industry state; and (v) that the actual advertising space purchased by each firm is given by (2). In practice, given that in equilibrium $p_{a,i,t}$ is such that total advertising expenditure in industry i exactly purchases α units of ad space, then, in all industries with more than one differentiated firms, $\alpha_{j,i,t}$ will be given by the second argument in (2), and so there will be an anticompetitive motive to advertise: by increasing $e_{j,i,t}$, firm j will achieve to increase the actual price for advertising space and thus reduce the advertising space of competitors, which will reduce the probability consumers learn about competitors. In industry states with only one differentiated firm, there is trivially no anticompetitive motive because there is no competitor and so the unique firm has no incentive to spend more than $p_{a,i,t}\alpha$, and so in such industry states $\alpha_{j,i,t}$ will be given by the first argument in (2).

Given that I focus on Markov perfect equilibrium, then the firm problem can be expressed in recursive form, with the value of the firm being a function of the state. Given that profits are linear on E_t, by guess and verify, the value of the firm is also linear in E_t. Therefore, defining $V_j(\mathcal{J}_{i,t}, \vec{M}_{i,t}) = \frac{V_{j,i,t}}{E_t}$, $\hat{e}_{j,i,t} = \frac{e_{j,i,t}}{E_t}$, $\hat{p}_{a,i,t} = \frac{p_{a,i,t}}{E_t}$ and $\pi_j(\omega_{j,i,t}, \mathcal{J}_{i,t}, \vec{M}_{i,t}) = \frac{\pi_{j,i,t}}{E_t}$ and using the Euler equation and that in the stationary equilibrium it will be $T_t = T$, we can write the dynamic firm problem recursively as

$$V_j\Big(\mathcal{J}, \vec{M}\Big) = \max_{\hat{e}_j} \Big\{\pi_j\Big(\omega_j, \mathcal{J}, \vec{M}\Big) - \hat{e}_j + \beta \mathbb{E} V_j\Big(\mathcal{J}', \vec{M}'\Big)\Big\}$$

$$\text{s.t. } \{\hat{e}_k\}_{k\in\mathcal{J}\setminus\{j\}},\ \{\mathcal{M}_k\}_{k\in\mathcal{J}},\ T,\ (9),\ (10),\ (14),\ (34),\ (35),\ (2)$$

We can decompose the FOC into the three motives to advertise: the informative motive (increase ρ_j), the anticompetitive motive (decrease $\rho_{j'}$, $j' \neq j$), and the persuasive motive (increase $s_{\mathcal{I},j,i}$ for $\mathcal{I} \in \mathcal{P}_j$):

$$1 = \underbrace{\frac{\partial \pi_{j,i}}{\partial e_j}}_{\substack{\text{Persuasive}\\ \text{motive}}} + \underbrace{\frac{\partial V_j}{\partial \rho_j}\frac{\partial \rho_j}{\partial \alpha_j}\frac{\partial \alpha_j}{\partial e_j}}_{\substack{\text{Informative}\\ \text{motive}}} + \underbrace{\sum_{j'\neq j}\left(-\frac{\partial V_j}{\partial \rho_{j'}}\right)\frac{\partial \rho_{j'}}{\partial \alpha_{j'}}\left(-\frac{\partial \alpha_{j'}}{\partial e_j}\right)}_{\text{Anticompetitive motive}} \tag{20}$$

In section 3.2 I show how the intensity of the three motives evolve with firm age.

2.2.4 Entrepreneurs problem

The entrepreneurs in an industry $(\mathcal{J}, \vec{M})$ choose N_e to maximise their expected value:

$$v^e(\mathcal{J}, \vec{M}) = \max_{N_e} \Big\{-N_e\hat{w} + \beta z_e \mathbb{E}_e V_e\Big(\mathcal{J}' \cup \{e\}, \vec{M}'\Big)\Big\},\ \text{s.t. } z_e = \phi_s N_e^{\frac{1}{2}}, \tag{21}$$

where $\mathbb{E}_e V_e\Big(\mathcal{J}'\cup\{e\}, \vec{M}'\Big)$ is the expected value of being a new firm conditional on successfully creating a new differentiated good (so, the expectation comes from the uncertainty on which of the $\mathcal{J}$ incumbents will survive). Then, the equilibrium labor employed in entry in an industry $(\mathcal{J}, \vec{M})$ will be:

$$N_{e,(\mathcal{J},\vec{M})} = \left(\frac{\phi_s}{2\hat{w}}\beta \mathbb{E}_e V_e\Big(\mathcal{J}' \cup \{e\}, \vec{M}'\Big)\right)^2 \tag{22}$$

2.2.5 Stationary distribution.

In the Appendix 6.10 I prove that, for any aggregates $\hat{w}$ and T given, with their associated solutions of the firms and entrepreneurs problems $\{\alpha_{j,(\mathcal{J},\vec{M})}, N_{e,(\mathcal{J},\vec{M})}\}$, the probability that an industry is at a given state $(\mathcal{J}, \vec{M})$ converges to an ergodic distribution (existence), which is independent of the initial state (uniqueness), and satisfies that the set of different states realised, call it Ω, is at most countably infinite.[16]

By Law of large numbers, this implies that the economy converges to a stationary distribution associated to the aggregates $\hat{w}, T$. Let $\mu_{(\mathcal{J},\vec{M})}$ be the mass of industries in state $(\mathcal{J}, \vec{M}) \in \Omega$ in this stationary distribution. If $\hat{w}, T$ are consistent with this stationary distribution, then we are in the stationary equilibrium.
Computationally, the stationary distribution is a complicated object, and the method used to obtain it, described in Appendix 6.11, is a computational contribution of this paper.
In Appendix **??**, I provide plots that help visualize the stationary distribution for the calibrated model.

2.2.6 Media sector problem.

Given the symmetry of media firms, consumers allocate their media time T equally among the media firms, and production firms allocate their advertising expenditure equally among the media firms. Therefore, all media firms have the same profits, and so each media firm has positive profits if and only if the overall profits in the media sector are positive. Then, since there is free entry into the media sector, profits in the media sector must be zero in equilibrium; so, the equilibrium Q_t satisfies:

$$\int_0^1 \sum_{j \in \mathcal{J}_{i,t}} \hat{e}_{j,i,t} E_t di + w_t \bar{N}_m - w_t \left(\frac{Q_t}{A}\right)^2 = 0 \tag{23}$$

where recall that $\bar{N}_m$ is the labor in media employed by the public sector. In the stationary equilibrium, $Q_t = Q$ is constant.

2.2.7 Labor market clearing.

The labor market must clear, that is, the amount of labor supplied has to be equal to the labor demanded by the production firms, media firms and entrepreneurs. Without any loss of generality (just a change in the units we measure labor), I can normalize labor supply N to 1.

$$1 = N = \int_0^1 \left(\sum_{j \in \{0\} \cup \mathcal{J}_{i,t}} N_{j,i,t} + N_{e,i,t} \right) di + N_{m,t} \tag{24}$$

where $N_{j,i,t} = \frac{s_{j,i,t}}{\mathcal{M}_{j,i,t}} \hat{w}^{-1}$, and $N_{e,i,t}$ and $N_{m,t}$ are given by (22) and (23), respectively. This pins down the equilibrium relative wage $\hat{w}_t$, and verifies that it is constant in the stationary

[16]Note that I have not formally proved whether the solution of the firms and entrepreneurs problems is unique. One could prove unicity by imposing restrictions on how the players make their decisions.

equilibrium.

2.2.8 Aggregate output and representative consumer conditional on age.

Define the aggregate consumption good as the geometric mean of the individuals' aggregate consumption goods; that is $\ln C_t = \int_0^1 \ln C_{\ell,t} d\ell$. Using the definitions of $C_{\ell,t}$ and $C_{\ell,i,t}$, together with $c_{\ell,j,i,t} = \frac{s_{\ell,j,i,t}}{\mathcal{M}_{j,i,t}} \hat{w}^{-1}$ and $c_{\ell,0,i,t} = s_{\ell,0,i,t} A_0 \hat{w}^{-1}$, and interchanging the integrals over ℓ and i, we obtain the level of the consumption good:

$$\ln C = -\ln \hat{w} + \sum_{(\mathcal{J},\vec{M})\in\Omega} \mu(\mathcal{J},\vec{M}) \sum_{\mathcal{I}\in\mathcal{P}(\mathcal{J})} M(\mathcal{I}) \frac{\sigma}{\sigma-1} \ln \left(\left(s_{\mathcal{I},0,(\mathcal{J},\vec{M})} A_0\right)^{\frac{\sigma-1}{\sigma}} + \sum_{j\in\mathcal{I}} \omega_{j,(\mathcal{J},\vec{M})} \left(\frac{s_{\mathcal{I},j,(\mathcal{J},\vec{M})}}{\mathcal{M}_{j,(\mathcal{J},\vec{M})}} \right)^{\frac{\sigma-1}{\sigma}} \right) \tag{25}$$

The aggregate price index of the economy is P_t such that $P_t C_t = E_t$, and it is the numeraire (i.e. $P_t = 1$). GDP in the economy is given by $Y = E + \sum_{(\mathcal{J},\vec{M})\in\Omega} \mu_{(\mathcal{J},\vec{M})} N_{e,(\mathcal{J},\vec{M})} \hat{w} E$.

Finally, note that applying a law of large numbers to the continuum of industries, two consumers with the same age will have the same level of aggregate consumption good. That is, although they will differ on their awareness sets for particular industries, at the aggregate level they will have the same distribution of awareness sets. This points to a potentially interesting extension where firms can target consumers based on the observable age.

3 Quantitative analysis

3.1 Calibration

In this section, I describe the calibration of the model, and the details of the data sources and how the moments are computed are provided in the Appendix 6.1. One of the main components of the model is firms' customer base accumulation, which has a strong relationship with firm size, both in the model and in the data, as pointed by the empirical literature cited in the introduction. Therefore, it is important that the model reproduces the evolution of the average firm sales growth by age, in order to calibrate this customer base building process. In particular, I target the constant and the linear coefficient from the fitted line of the plot of average firm relative sales growth by age. Also, as shown in section 3.2, the intensity of the different motives to advertise varies with firm size, so the coefficient from a regression of advertising expenditure and sales is a good candidate to discipline the model. To compute these three moments I use Compustat data. Given that firms typically enter Compustat a few years after their foundation (and certainly not with zero customers as it is assumed in the model for new firms), for the computation of the model-implied moments of these three targets, I assume that firms in the model are unobserved until they are at least five years old.

I estimate the model for the US at an annual frequency and set the consumer discount rate to $\beta = 0.98$. I also set (i) $\delta = 0.01$ corresponding to the mortality rate of 1% in the data, (ii) the concavity parameter for the persuasive advertising $\nu_c = 0.2972$ is taken from (the inverse) Cavenaile et al. (2024a), (iii) given that public sector spending on media represents roughly 0.008% of US GDP, dividing this by the (capital-adjusted) labor share, I set $\bar{N}_m = \frac{8 \cdot 10^{-5}}{0.8359}$, and (iv) I set $\kappa = 0.1151$ corresponding to the entry rate in the data. Acknowledging the difficulty to find good proxies for the utility value of media goods, I leave the weight of the entertainment good on the utility function, υ, uncalibrated and all the exercises involving welfare are made for a range of values of υ. This leaves 8 moments to estimate: the elasticity of substitution parameter, σ; the relative productivity of the small firms producing the generic goods, A_0; the scale parameter for the persuasive effect of advertising, ν_s; the scale and convexity parameters for the informative effect of advertising, (ψ_s, ψ_c); the exogenous learning probability, $\hat{\rho}$; the scale parameter regulating the creation of new products, ϕ; and the aggregate productivity of the media sector, A. Apart from A, which can be derived directly from 23 using the target values for aggregate advertising expenditure and labor shares and the fraction of time in media, the rest of the 7 parameters are estimated jointly through a Simulated Method of Moments estimation procedure. Apart from the three moments described above concerning the average firm growth by age and the relationship between advertising and sales, at the aggregate level, I target the sales-weighted average markup and standard deviation, the aggregate advertising expenditure as a percentage of GDP, the fraction of time spent in media, and the labor share. Given that there is no physical capital in the model, for comparability, I take the labor share as the share of labor income among labor income and profits, following Cavenaile et al. (2024a). Table 1 summarises the results of the calibration. Panel A reports the parameter values, while Panel B reports both the model-implied moments and the empirical ones. The model does well in matching the moments. In addition to the targeted moments, the calibrated model also features an inverted-U relationship between advertising expenditure and relative sales as documented in Cavenaile et al. (2024a).

Note that Compustat is not the ideal dataset to discipline the growth pattern of firms in the model for the following reasons. First, firms do not automatically enter Compustat when they are born, and they may enter at different stages of the life cycle. Second, contrary to the model, firms may grow by expanding to new geographical markets or new product lines. Figure 1 plots the average firm relative sales growth rate both in the model and in the data. Note that in the model, if a firm had a constant $\rho_{j,i}$ (this is the case of a firm that has always

Table 1: Parameter values and targeted moments

A. Parameters

	Parameter	Description	Calibration	Value
Preferences	β	Discount rate	External	0.98
	σ	CES consumption	Internal	5.0625
	υ	weight of leisure	Uncalibrated	-
Persuasive	ν_s	Scale parameter	Internal	0.1250
	ν_c	Convexity parameter	External	0.2972
Learning	ψ_s	Scale parameter	Internal	0.2194
	ψ_c	Convexity parameter	Internal	0.4500
	$\hat{\rho}$	Exogenous learning	Internal	0.1000
	δ	Mortality rate	External	0.01
Media sector	A	productivity media firms	Internal	3.2087
	$\bar{N}_m$	public sector media	External	$9.5705 \cdot 10^{-5}$
Generic good	A_0	Productivity	Internal	0.5047
Entry/Exit	κ	Exit rate	External	0.1151
	ϕ	Entry scale	Internal	0.6422

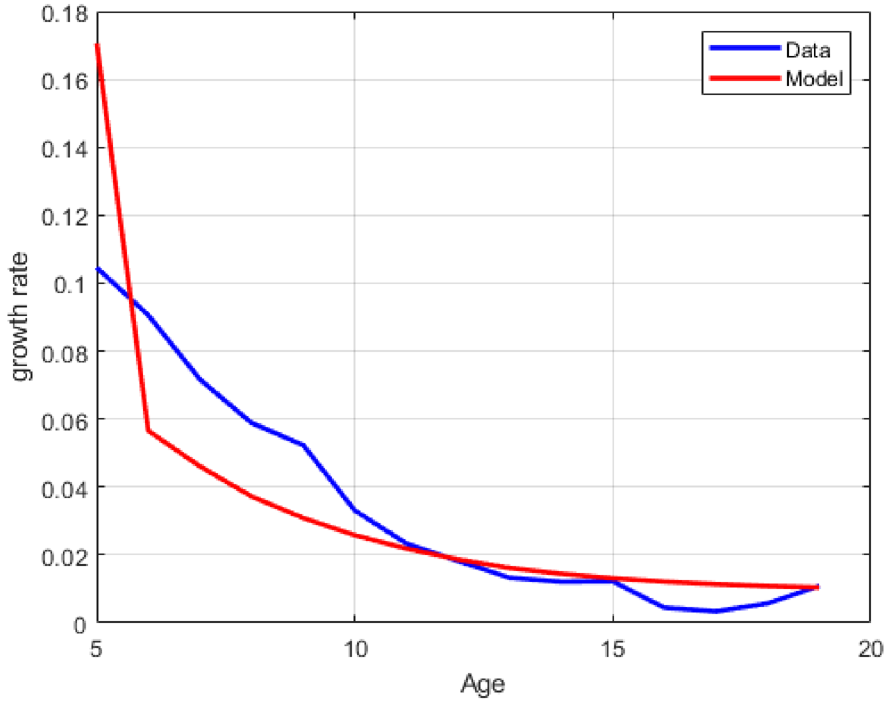
B. Moments

Moment	Data	Model
Sales-weighted average markup	1.3498	1.3281
St.Dev. Markup	0.3460	0.3479
Labor share (capital-adjusted)	0.8359	0.8392
Advertising/GDP	2.2%	2.1535%
Fraction of time in media	0.552	0.552
Intercept (firm growth, age)	0.0784	0.0831
Linear(firm growth, age)	-0.0061	-0.0063
Linear(adv. exp, market share)	0.6710	0.6501

Notes. Panel A reports the parameter values. Panel B reports the simulated and empirical moments. Details on data sources and how this moments are computed can be found in the Appendix 6.1

been the single differentiated firm of the industry), then growth would be monotonically decreasing, pushed by a mechanical force: given that the population is constant, as the firm's customer base expands, the growth rate slows down because (i) a given increase in customers has a smaller relative impact, and (ii) there are fewer non-customers remaining. Things get noisier when there are other competitors and there is entry and exit.

Figure 1: Average firm growth by age



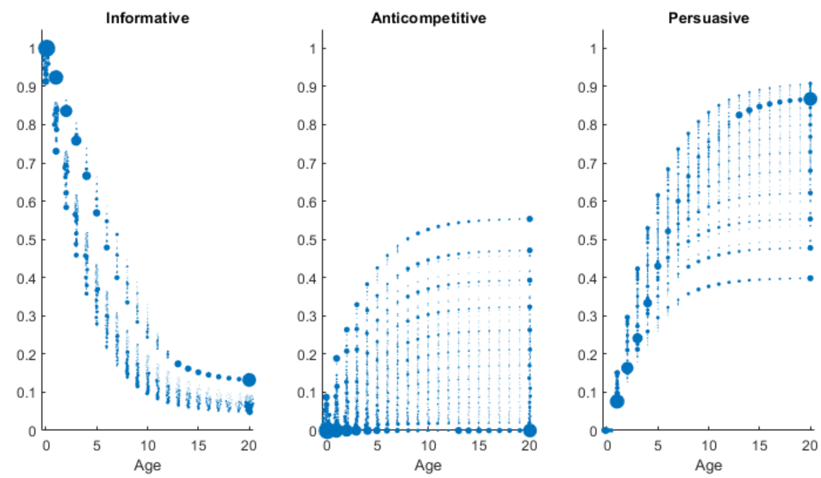
Notes. This figure displays the average firm relative sales growth by age both in the data (blue) and in the model (red). Given that firms typically enter Compustat a few years after their foundation, for comparability I assume age 0 in Compustat corresponds to age 5 in the model.

3.2 Advertising motives and firm age and size

In this section, I quantify the share of the incentives to advertise attributable to each of the three motives and examine their relationship with firm age and size. The intuition is clear: smaller or younger firms—those that are unknown to most consumers—have more potential customers to acquire. In the extreme case, a firm known by all consumers would have no incentive to advertise for informational purposes. Conversely, the anticompetitive motive, which is about retaining market power over the current customers by reducing the probability that they learn about competitors, becomes stronger as the customer base grows. A firm that is unknown to all consumers also has some incentive to prevent them from learning about other firms (since it internalizes that these consumers may eventually become customers, and thus wants them to know as few goods as possible), but intuitively, the incentive to prevent a consumer from learning about a competitor is higher when the consumer is already a current customer rather than just a potential one. Finally, the persuasive motive also tends to be bigger in older or larger firms. Intuitively, if advertising persuades current customers to spend more, the increase in revenue will be larger if there are more customers.

The decomposition of the FOC of advertising expenditure in (20) allows us to see the share of the firm's marginal value of advertising coming from each of the three motives. Using this observation, Figure 2 displays the share of the marginal value of advertising attributable to each of the three motives for all the firms in the stationary equilibrium. Three observations

Figure 2: Marginal intensity of the advertising motives by age



Notes. This figure displays the values from the terms of the FOC corresponding to the informative motive (left panel), the anticompetitive motive (middle panel) and the persuasive motive (right panel), for all the firms in the stationary equilibrium, where the relative size of each dot indicates the share of this firm type in the stationary distribution.

can be drawn. First, that there is significant heterogeneity, which indicates that age is far from being a sufficient statistic. This shows that industry dynamics play a key role (i.e. competition matters). Second, despite the variability, it can be observed that the informative motive is negatively associated with age, while the anticompetitive and the persuasive motives are positively associated with it. Finally, firm age appears to play a particularly important role in distinguishing the motives during the first 5–10 years of a firm's life.

By aggregating the previous shares, weighted by total industry advertising expenditure, we obtain an indicative breakdown of the advertising expenditure attributable to each motive. This exercise suggests that 57.81% of the incentives correspond to the informative motive, while 33.74% correspond to the persuasive motive and 8.45% to the anticompetitive motive. Figure 10 further repeats this decomposition exercise conditional on the number of firms in the industry. This reveals that the persuasive motive is more important as the number of competitors increases.

Because of the positive link between firm age and size (either sales or customer base, see

Figure 11), we obtain qualitatively similar plots when firm size is used instead of age. This numerical result is further supported by the following analytical result:

Proposition 1 *If distribution $\vec{M}_2$ is obtained from $\vec{M}_1$ by adding $\{j\}$ to the awareness sets of some consumers (that is, formally, if $\vec{M}_1$, $\vec{M}_2$ satisfy $M_2(\mathcal{I}\cup\{j\})-M_1(\mathcal{I}\cup\{j\}) = M_1(\mathcal{I})-M_2(\mathcal{I}) \geq 0$ for every $\mathcal{I} \in \mathcal{P}_{-j} = \{\mathcal{I} \in \mathcal{P} | j \notin \mathcal{I}\}$), then (for now, the proof is keeping the advertising choices fixed):*

1. *Firm j's informative motive is smaller in $\vec{M}_2$. That is, the informative motive is stronger in smaller firms.*

2. *Firm j's anticompetitive and persuasive motives are bigger in $\vec{M}_2$. That is, both the anticompetitive and the persuasive motives are stronger in bigger firms.*

Proof. See the Appendix 6.7 ■

3.3 Counterfactuals shutting motives

How do each of the three motives affect the aggregates? Is the anticompetitive motive necessarily bad? What are the aggregate effects of shutting down advertising? This section addresses these questions. To do so, I compare the baseline economy with three counterfactual scenarios. The first is an economy where firms neglect the anticompetitive motive; that is, they don't internalise that by increasing their advertising expenditure they are effectively reducing the amount of consumers who learn about competitors. To be precise, this is done by removing the anticompetitive component from the firm's first order condition. In the second counterfactual, in addition, firms also neglect the persuasive motive; that is, they don't internalise that advertising increases current customers' spending. In the third one, the informative motive is also shut down, meaning firms don't advertise at all, and so consumers only learn through the exogenous probability; i.e., $\rho_{j,i,t} = \hat{\rho}$. This exercise illustrates what the economy would look like if firms neglected some of the motives to advertise. Such negligence alters firms' decisions, which in turn also has general equilibrium consequences.

Table 2 reports some relevant statistics for the counterfactuals and the benchmark. The second row shows the level of the consumption good assuming that the persuasive advertising is deceptive (i.e., consumers make their purchasing decisions based on $\omega_{j,i,t}$, but then they derive utility as if $\omega_{j,i,t} = 1$).

First, as intuition suggests, without the anticompetitive motive, smaller firms face less competition for advertising space, allowing them to grow faster, which increases competition and consequently lowers markups. Additionally, improved growth prospects increase entrepreneurs' incentives to create new products, driving up the entry rate. However, these

Table 2: Comparison of counterfactuals with firms neglecting the anticompetitive and/or the persuasive motives

	Benchmark	No Anticompetitive	Only Informative	No motive
C	0.7260	0.7268	0.7226	0.6049
C no taste shifter	0.6898	0.6916	0.6935	0.6049
Q	0.5150	0.4860	0.3536	0.0314
Adv/GDP	2.1535	1.9137	1.0054	0
w	0.8392	0.8376	0.8345	0.9010
Avg Number of Firms	1.4148	1.4233	1.4406	1.1087
Sales-Weighted Average Markup	1.3281	1.3260	1.3135	1.1698
Coefficient advertising vs market share	0.6501	-2.5961	-15.0744	.

Notes. In the 'No Anticompetitive' counterfactual, firms make their decisions neglecting the anticompetitive motive; in the 'Only Informative' counterfactual, firms neglect both the anticompetitive and the persuasive motives; and in the 'No motive' firms neglect all the incentives to advertise, so they don't advertise and there is only the exogenous learning, $\rho_{j,i,t} = \hat{\rho}$.

positive effects on C are mitigated by a negative general equilibrium effect. Removing one incentive to advertise decreases the demand for advertising, which in turn reduces aggregate advertising expenditure, leading to a lower supply of media goods. As a result, consumers spend less time on media, meaning they are less exposed to advertising, which renders advertising less effective. This explains the negligible overall effect on C.[17] Therefore, although the anticompetitive motive has an overall negative effect on consumption, it is not necessarily the case that welfare would be higher in a counterfactual economy without it, due to its contribution on the provision of media goods.

The counterfactual where, in addition, the persuasive motive is shut down suggests that the persuasive motive has an overall positive effect on consumption. The second line shows that this positive effect is due to consumers enjoying the advertised goods more. Similarly to the anticompetitive motive, shutting down the persuasive motive allows smaller firms to capture a larger share of the advertising space, which increases entry and lowers markups. Moreover, as in the first counterfactual, although shutting down the motive has a significant effect on firms' advertising decisions (in this case, advertising expenditure falls by 47.46%), its effect on C is very modest. Thus, on the aggregate, both motives matter mostly through media goods.

Finally, the last counterfactual shows that shutting down advertising would decrease consumption by 16.68% relative to the benchmark. In this counterfactual, the media sector only receives revenue from the public sector. In this counterfactual, firms' customer base only grows via the exogenous learning. This implies that, in the equilibrium, the differentiated

[17] Relatedly, in the Conclusion I discuss that the current specification of $\rho_{j,i,t}$ may exhibit excessive diminishing returns.

firms will tend to be smaller, which decreases the incentives to enter. Given that there are fewer differentiated firms and that they are, on average, smaller in size, the generic goods capture a larger market share, which drives the average markup down.

4 Welfare: Planner Problem and Taxation

In this section, I first identify the sources of inefficiency in the model and then solve the informationally-constraint planner problem and compare it to the decentralized equilibrium. Finally, I examine taxation on advertising.

4.1 Social planner problem

The model is inefficient for several reasons. First, the dispersion of markups typical in oligopolistic setups leads to labor misallocation in the production sector. Second, when choosing their advertising expenditure, firms do not internalize the entertainment value of media goods, which are financed through advertising. This points to an underprovision of media goods, as in Greenwood et al. (2024). Additionally, there are three sources of inefficiency coming from the advertising choices that are characteristic of the current paper: the anticompetitive motive, the lack of full appropriability, and business-stealing. Note that we can distinguish between inefficiencies in the level of advertising expenditure (or, equivalently, in the prices of advertising space or in the provision of media goods) and inefficiencies in the allocation of advertising space. In this sense, the anticompetitive motive points to too much advertising and shifts the allocation of advertising space toward older firms. The lack of full appropriability, meaning that firms cannot extract the full surplus, pushes towards having too little advertising. Finally, business-stealing here refers to firms not internalizing the losses from the reduction in the consumption of other goods when consumers learn about their product, which pushes toward excessive advertising, especially in industries with more competitors. Moreover, there is also inefficient entry, again due to lack of appropriability and business-stealing.

To assess the importance of these inefficiencies, I solve the following planner problem and compare the resulting equilibrium with the decentralized one.[18] The planner has full control over production, media, and entrepreneurial decisions but cannot affect consumers' behavior; that is, the learning process and the choices regarding consumption and media time remain as they are in the decentralized equilibrium. Its goal is to maximize aggregate utility, with

[18]This is work in progress. Given that there are numerous sources of inefficiency, a more insightful exercise would be to compare allocations where the planner takes control of one additional decision.

all individuals weighted equally. Formally, the planner solves:

$$\max_{\{N_{M,t},N_{j,i,t},N_{e,i,t},p_{j,i,t},\alpha_{j,i,t}\}} U = \sum_{t=0}^{\infty} \beta^t \int_0^1 [\ln C_{\ell t} + L_{\ell t}]\, d\ell$$

s.t. $C_{\ell,t}$ from (5), $C_{\ell,i,t}$ from (6), $c_{\ell,j,i,t}$ from (11), and $L_{\ell,t}$ from 4, with $T_t = Q_t$ (Consumer choices)

$$y_{j,i,t} = N_{j,i,t}, \quad y_{0,i,t} = A_0 N_{0,i,t}, \quad Q_t = A N_{m,t}^{\frac{1}{2}} \quad \text{(Production functions)}$$

$$1 = N_{m,t} + \int_0^1 \left(\sum_{j\in\{0\}\cup\mathcal{J}_{i,t}} N_{j,i,t} + N_{e,i,t} \right) di \quad , \quad w_t = E_t \quad \text{(Resource constraints)}$$

$$\sum_{j\in\mathcal{J}_{i,t}} \alpha_{j,i,t} = \alpha, (9), (10), (14), (34), (35) \quad \text{(Learning process)}$$

$$z_{e,i,t} = \phi N_{e,i,t}^{\frac{1}{2}}, (34), (35) \quad \text{(Entry and exit)}$$

I leave the details of the solution in the Appendix 6.9. The planner sets prices equal to marginal cost times a markup (or a tax) that allows the planner to pay for the labor to produce the media goods and for entry. That is, $p_{j,i,t} = \tau w_t / A_j$, with $\tau = E_t/(w_t N_t^P)$, where N_t^P is the labor used in the production sector.

For the dynamic problem of advertising and media, as in the baseline model, I focus on the stationary Markov perfect equilibrium. The social planner has to decide on (i) how to allocate the ad space among the differentiated firms of each industry, $\alpha_{j,i,t}$, (ii) how much labor to allocate to the media sector, $N_{m,t}$, and (iii) how much labor to allocate to creating new products in each sector, $N_{e,i,t}$.

First, let's see the social planner choice of $\alpha_{j,i,t}$. The allocation of the ad space has to be such that the marginal social gain of increasing the ad space given to each firm is the same, since otherwise we could improve the allocation. Formally, letting $\ln C_{i,t} = \int_0^1 \ln C_{\ell,i,t} d\ell$ be the total consumption good of industry i, and $U_X = \sum_{t=0}^{\infty} \beta^t \mathbb{E} \ln C_{i,t}$ be the expected life-time utility derived from an industry whose current state is X; it must be

$$\frac{\partial \ln C_{X,t}}{\partial \alpha_{j,X}} + \beta \frac{\partial \mathbb{E} U_{X'}}{\partial \rho_{j,X}} \frac{\partial \rho_{j,X}}{\partial \alpha_{j,X}} = \hat{h}_X \quad \text{for some } \hat{h}_X \text{ and all } j \in \mathcal{J}_X, \text{ together with } \sum_{j\in\mathcal{J}_X} \alpha_{j,X} = \alpha \tag{26}$$

Note that the anticompetitive motive plays no role in the social planner's allocation of $\alpha_{j,X}$, as the planner directly chooses the ad space occupied by each firm. So, in deciding whether to give more ad space to one firm over another, the planner only considers the utility gains from informing more consumers and from enhancing customers' taste for that good.

Second, let's see the social planner choice of N_m. The planner takes into account that by employing more labor in media it will increase the aggregate quality Q of media, which has two effects: (i) it increases the level of entertainment L; and, by increasing the time spent in media, (ii) it increases the consumption good by increasing the probability of learning goods. The optimal N_m is given by

$$N_m = \frac{N^P}{2} \left(\upsilon Q^2 + \sum_{X\in\Omega} \mu(X) \hat{h}_X \alpha \right) \tag{27}$$

Note that, unlike existing literature, here the planner values the provision of media goods even if their entertainment value was negligible (i.e. even if $\upsilon = 0$), due to their role as a vehicle for spreading product awareness. Finally, the labor employed in entry in each industry satisfies:

$$N_{e,X} = \left(\frac{\phi N^P}{2}\beta\left(\mathbb{E}_e U_{X'} - \mathbb{E}_{-e} U_{X'}\right)\right)^2 \tag{28}$$

where $\mathbb{E}_e U_{X'}$ (resp. $\mathbb{E}_{-e} U_{X'}$) is the expected industry-utility conditional on successfully creating (resp. not creating) a new differentiated good (so the undertainty comes from the probabilities the incumbents exit). The relative wage is $\hat{w} = 1$ as consumers spend all the income they receive, which is w. The labor market clearing, using 49 and 27 pins down N^P:

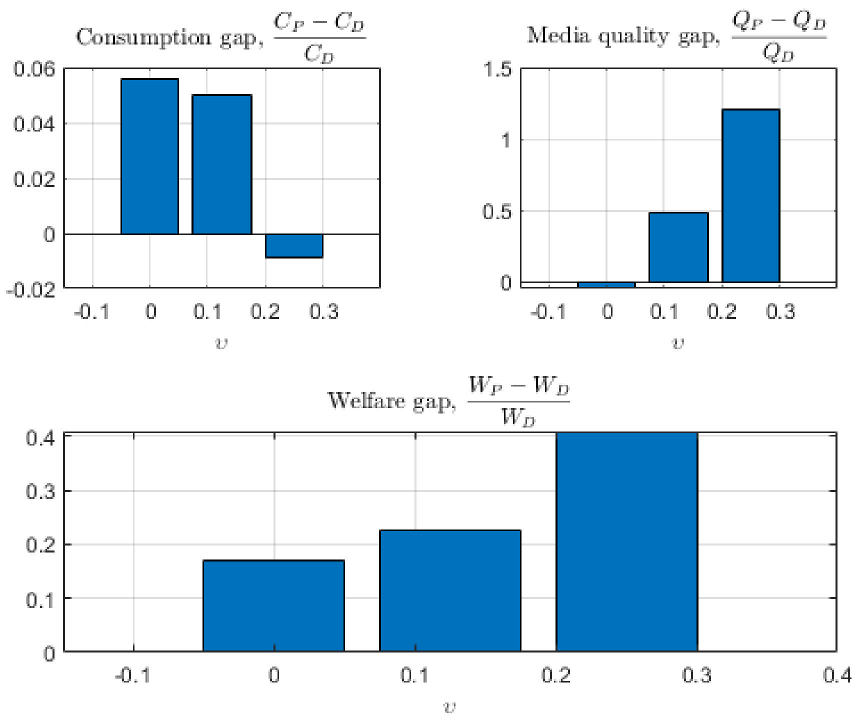
$$1 = N^P + N_e + N_m \tag{29}$$

Figure 3 compares the planner economy with the decentralized one for different values of the relative utility weight of the entertainment good, υ. As expected, as υ increases, the planner puts more weight on producing media goods, at the expense of consumption, which eventually is lower than in the decentralized equilibrium. More interestingly, when $\upsilon \to 0$ (that is, when spending time on media doesn't provide any direct utility gain to consumers), the supply of media qoods is larger in the decentralized equilibrium, which seems to indicate that, when $\upsilon \to 0$, there is too much advertising. However, it is important to remind that there are inefficiencies both in the level of advertising expenditure as well as in the allocation of the advertising space. Therefore, this doesn't mean that welfare in the decentralized equilibrium would improve if all firms reduced their advertising expenditure proportionally to emulate the same Q as in the planner's equilibrium. In other words, the inefficiencies from the misallocation in the advertising space may be mitigated with the 'overprovision' of media, through its effect on learning.

4.2 Taxing advertising: Uniform tax

Given that the decentralized equilibrium is inefficient, this section explores the welfare gains from the uniform tax on advertising that maximizes welfare. Here, in addition to $e_{j,i,t}$, firms pay $\tau_a e_{j,i,t}$ as taxes to the government, which are distributed as transfers to consumers. Figure 4 depicts the effect of this tax on final output, the quality of media, and welfare for different values of υ. As expected, the higher the entertainment value of media, the more valuable a subsidy on advertising becomes, as it magnifies the inefficiency arising from the fact that firms don't internalize the entertainment value of media goods. More interestingly, recall that we have seen that the decentralised equilibrium supplies more media than the

Figure 3: Welfare comparison of the planner and decentralised allocations



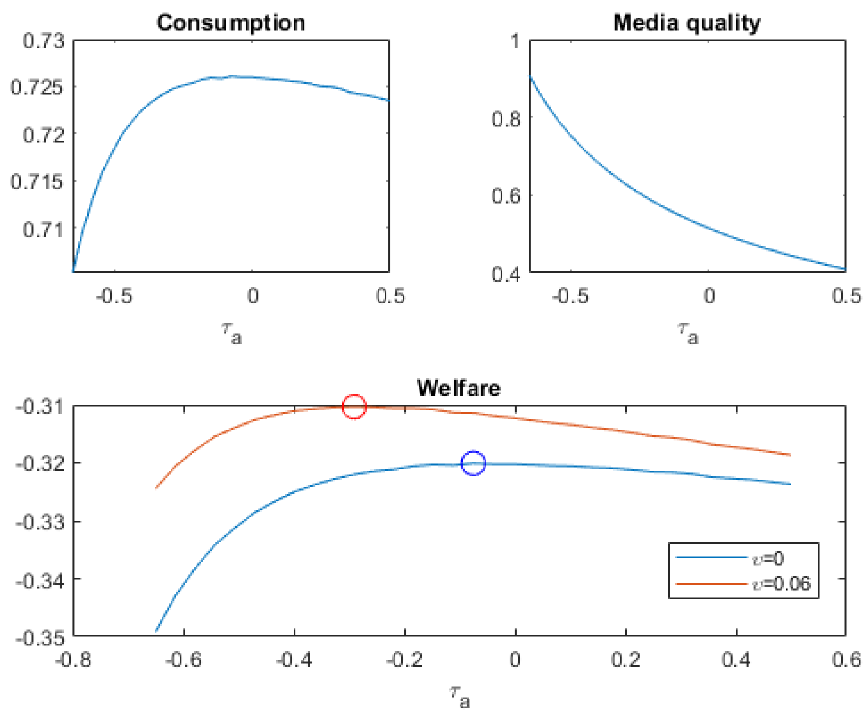
Notes. This figure displays the difference in final output (upper-left panel), in media time (upper-right panel), and welfare (bottom panel) between the planner's equilibrium and the decentralized one, relative to the decentralized one.

planner's equilibrium in the case of $\upsilon \to 0$, which seems to point to an overprovision of media. Actually, it turns out that even when $\upsilon \to 0$, the optimal tax is a subsidy. This suggests that the 'overprovision' of media, via increasing the time spent on media and thus the effectiveness of advertising, mitigates the inefficiencies from the misallocation of the advertising space. However, the welfare gains from such subsidy are very modest. This is despite the fact that increasing the tax significantly reduces time on media and that in section 3.3 we have seen that shutting down advertising has a sizable effect on C. The reason these findings are compatible with the negligible effect of the tax in the model is due to the diminishing returns in $\rho_{j,i,t}$; that is, setting T to 0 would have sizable effects, but a partial reduction of T has a small effect on the probabilities $\rho_{j,i,t}$.

4.3 Age-dependent tax

The observation in section 3.2 that the informative motive is decreasing with age, whereas the persuasive and anticompetitive motives are increasing with age, together with the fact that they have different welfare implications, leads us to think that we may achieve significant

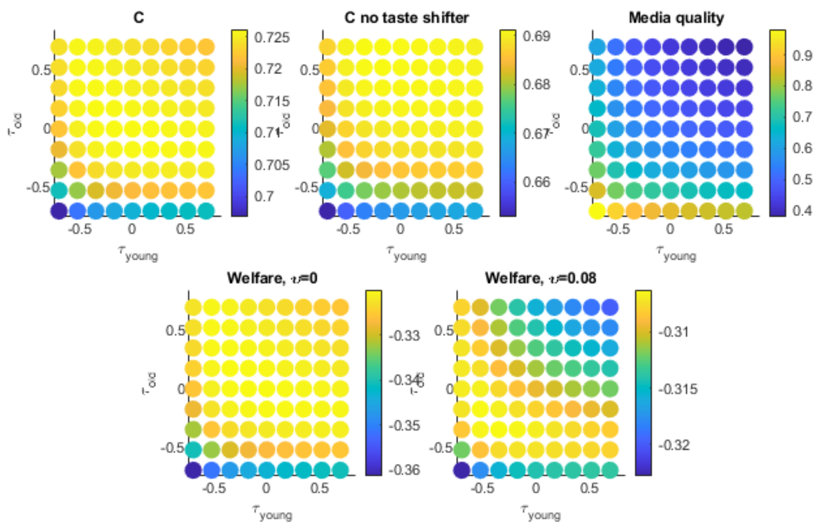
Figure 4: Welfare under uniform tax on advertising



welfare improvements by considering an age-dependent tax rather than applying the same tax to all firms. Assume now that firms pay τ_Y if their age is less than the age cutoff $\bar{a}$, and τ_O if their age is greater than or equal to $\bar{a}$. In particular, for this exercise, I have set $\bar{a} = 3$; this means that firms receive different tax treatment during their first three periods of life compared to afterwards.

This differential policy treatment makes the vector of ages of the firms an additional state. Note that firms that are at least $\bar{a}$ years old are indistinguishable by age; if all firms are older, then the firm problem is identical to the baseline with a uniform tax. However, for $a_j < \bar{a}$, we need to keep track of the particular age a_j; i.e., how close a firm is to $\bar{a}$ makes a difference. So, if $(a_1, \ldots, a_J)$ is the vector of ages (from older to younger), then the relevant age state is $\vec{a} = (\hat{a}_1, \ldots, \hat{a}_J)$, where $\hat{a}_j = \min(a_j, \bar{a})$.
Figure 5 illustrates the effect of this age-dependent advertising tax on (i) consumption, (ii) consumption if we assume that the persuasive effect of advertising is deceptive (as described in 3.3), (iii) media quality, and (iv) welfare for two values of υ. However, as with the uniform tax exercise, in the current version of the model, the gains from such a policy are minimal.

Figure 5: Welfare under age-dependent tax on advertising



5 Concluding remarks

The informative and persuasive aspects of advertising are widely accepted: firms may advertise to mitigate some information friction by informing consumers, but also to shift market shares from one firm to another. On top of this, I highlight that the fact that consumers' attention is limited introduces a novel motive to advertise: firms may want to divert consumers' attention away from competitors.

The two main contributions of this paper are, on the one hand, to develop a novel model that accommodates these three motives in a single framework, and, on the other hand, to solve a dynamic model of informative advertising where asymmetric firms play strategically in a dynamic game. Thus, the model allows us to think about how firms build their customer capital and how they interact with competitors' customer capital.

I first use the model to examine how the motives evolve along the firm's life-cycle, their contribution to total advertising, and their aggregate effects. The informative motive, which accounts for around half of the incentives to advertise, is stronger for younger firms, as these are less known. The persuasive and anticompetitive motives, which are stronger in older firms, while relevant from the firms' perspective, have an almost negligible effect on aggregate consumption. Instead, they mostly matter through the provision of entertaining media goods, as they significantly contribute to total advertising expenditure. However, completely shutting down advertising leads to a considerable 16.68% reduction in consumption Finally, given that there are several sources of inefficiency in the model, I compare the planner's allocation with the decentralized one, and study the welfare gains from taxing advertising.

I find that advertising should be subsidized, although the gains are small.

Two considerations are relevant in understanding the modest effect on aggregate consumption both when only some motives are shut down and when advertising is taxed. First, as explained in section 3.3, there is a general equilibrium effect that mitigates their impact. Second, the model features inherent diminishing returns to advertising: as more consumers become informed, fewer consumers remain to be informed. Therefore, the assumption that the probability of learning also exhibits diminishing returns may imply that the diminishing returns to advertising are too strong, so changes in the exposure to advertising have little impact on learning. This is work in progress, and my plan is to see how the results change using a different specification for the learning process.

References

Argente, D., Fitzgerald, D., Moreira, S., & Priolo, A. (2023). How do entrants build market share? the role of demand frictions. *American Economic Review: Insights*, *forthcoming*.

Bagwell, K. (2007). *The economic analysis of advertising*. Elsevier.

Cavenaile, L., Celik, M., Perla, J., & Roldan-Blanco, P. (2024b). A theory of dynamic product awareness and targeted advertising. *Journal of Political Economy*, *Revise and resubmit*.

Cavenaile, L., Celik, M., Roldan-Blanco, P., & Tian, X. (2024a). Style over substance? advertising, innovation and endogenous market structure. *Journal of Monetary Economics*.

Dinlersoz, E. M., & Yorukoglu, M. (2012). Information and industry dynamics. *American Economic Review*, *102 (2)*, 884–913. https://www.aeaweb.org/articles?id=10.1257/aer.102.2.884

Einav, L., Klenow, P. J., Levin, J. D., & Murciano-Goroff, R. (2022). Customers and retail growth. *Review of Economic Studies*, *Revise and resubmit*.

Foster, L., Haltiwanger, J., & Syverson, C. (2016). The slow growth of new plants: Learning about demand? *Economica*, *83*(329), 91–129.

Gourio, F., & Rudanko, L. (2014). Customer capital. *The Review of Economic Studies*, *81 (3)*, 1102–1136. https://doi.org/10.1093/restud/rdu007

Greenwood, J., Ma, Y., & Yorukoglu, M. (2024). "you will:" a macroeconomic analysis of digital advertising. *The Review of Economic Studies*, *forthcoming*.

Ignaszak, M., & Sedlacek, P. (2023). Customer acquisition, business dynamism and aggregate and growth. *The Review of Economic Studies*, *Revise and resubmit*.

Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *84(3)*, 1153–1176.

Molinari, B., & Turino, F. (2018). Advertising and aggregate consumption: A bayesian dsge assessment. *The Economic Journal*, *128(613)*, 2106–2130.

Rachel, L. (2024). Leisure-enhancing technological change. *mimeo*.

6 Appendix

6.1 Calibration Appendix: Data sources and Computation of moments

1. **Sales-weighted average markup, sales-weighted standard deviation of markups, labor share, entry rate, and aggregate advertising expenditure as a percentage of GDP.** Taken from Cavenaile et al. (2024a). Following Cavenaile et al. (2024a), given that there is no physical capital in the model, I target the labor share among labor income and profits. Given that $\frac{wL}{wL+\pi+rK} = \frac{wL}{wL+\pi}\frac{wL+\pi}{wL+\pi+rK} = \frac{wL}{wL+\pi}\left(1 - \frac{rK}{wL+\pi+rK}\right)$; then, the target used is obtained from dividing the labor share by one minus the capital share. In the model, given that labor supply is normalised to 1, then labor share equals w.
   The entry rate in the model is the average number of new firms (i.e. the average probability of creating a new product).

2. **Fraction of time in media.** According to Statista, people in the US spend on average 751 minutes per day in media, which corresponds to the 0.521528 of time. Note that in this measure of media time multitasking is counted separately; that is: it counts the time spend in media while also doing other activities (e.g. commuting to work, breaks at work, listening a podcast while cooking or running), and duplicated media time when using multiple forms of media simultaneously (e.g. watching the TV while using a phone will count double).

3. **Coefficient of a regression of advertising expenditure on relative sales.** This and the growth by age moments are computed using Compustat data for the time period 1976-2018. Both in the model and in the data, I take the logarithm of advertising expenditure and then I standardise it by subtracting their means and dividing by their standard deviation for comparability. In the data, I regress the standardised logarithm of advertising expenses on relative sales of the firm in its SIC4 industry, controlling for the same set of controls used in Cavenaile et al., namely: profitability, leverage, market-to-book ratio, log R&D stock, firm age, the coefficient of variation of the firm's stock price, the number of firms in the industry, and a full set of year and SIC4 industry fixed effects. In the model, I regress the standardised logarithm of advertising expenses, $p_{a,i,t}e_{j,i,t}$, on market shares, $s_{j,i,t}$, with industry fixed effects. Table 3 shows the results of the empirical regression:

4. **Constant and slope of the fitted line of average firm relative sales growth**

Table 3: Advertising and relative sales in the data

	log advertising expenses
Relative sales	0.671 (0.0448)***
R^2	0.6056
N	40,007

Notes. Robust asymptotic standard errors (in parenthesis) are clustered at the firm level. The sample period is from 1976 to 2018. The regression controls for profitability, leverage, market-to-book ratio, log R&D stock, firm age, the coefficient of variation of the firm's stock price, the number of firms in the industry, and a full set of year and SIC4 industry fixed effects.

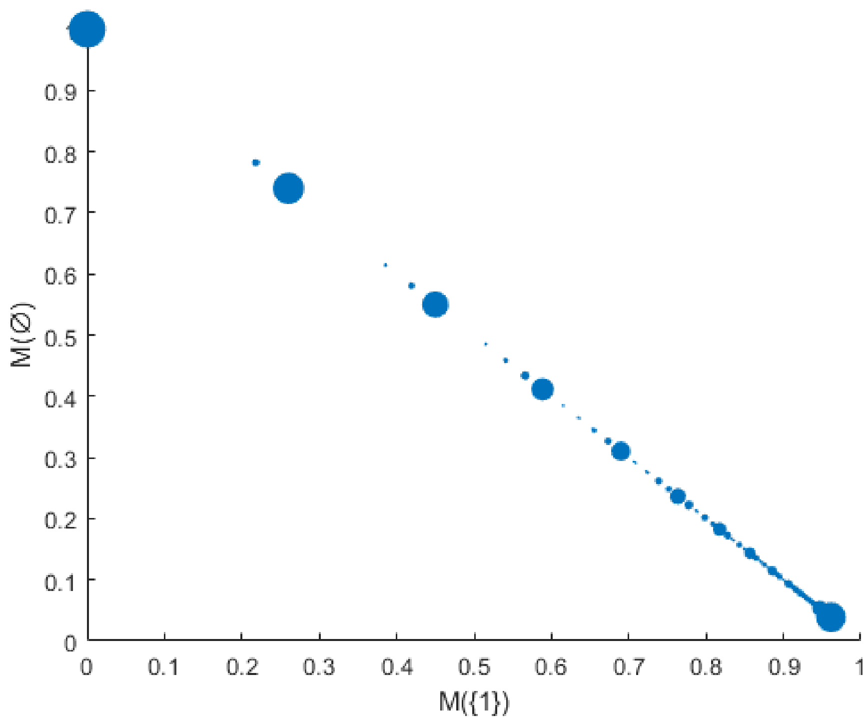
by age. In Compustat, I define age as the number of years since the first appearance of the firm in Compustat. First, for comparability with the model, where there is no aggregate growth, I compute growth rates of relative sales of the firm in its SIC4 industry. Second, firms in the data may experience big jumps on sales through expansion to new markets or via mergers and acquisitions, and I am interested in the average evolution of firm growth in the absence of such disruptive events; therefore, I drop all the observations of a firm posterior to a big change in their relative sales. In particular, if a firm's relative sales increase by more than 100% or decrease by more than 50%, this observation and the posterior ones of this firm are dropped. Then, I take the average firm relative sales growth grouping all the observations with the same age. In the model, I redefine age by subtracting 5 (as I am assuming that age 5 in the model corresponds to age 0 in Compustat). Given the average firm relative sales growth by age, $\bar{g}_a$, I define the fitted line $\hat{g}_a = \beta_0 + \beta_1 a$, where a is age. The coefficients β_0 and β_1 are the targeted moments.

5. **Calibration of the public sector financed media $\bar{N}_m$.** According to the US Government Accountability Office, the federal government spent \$14.9 billion over the last 10 fiscal years (2014-2023). Then, I use that federal governments spent roughly \$1.49 billion per year. In addition, federal appropriations for CPB (Corporation for Public Broadcasting) amounted to \$477 million in fiscal year 2023. So, the estimate I use for public sector spending on media is (\$1.49 + \$0.477) billion, which I divide for the US GDP in 2023, \$27360 billion. This gives 0.008% of GDP, which divided by $w = 0.8359$ gives the $\bar{N}_m = 9.5705 \cdot 10^5$.

6.2 Visualizing the stationary distribution

Figure 6 depicts the industries with only one differentiated good. Recall that for such industries industry state is fully characterized by the mass of consumers that know the only differentiated good. For industries with two differentiated goods, the industry space is 3-dimensional. Figure 7 depicts such industries in the stationary distribution.

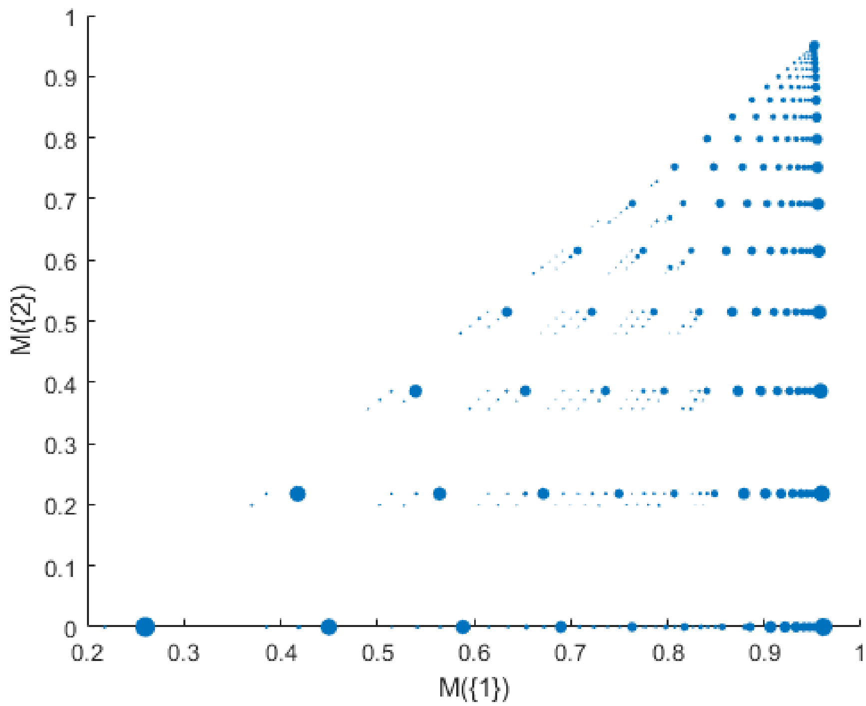
Figure 6: Industry states in industries with one firm



Notes. This figure displays the industry states in industries with only one differentiated firm. The size of the dots is proportional to the frequency of the industry state in the stationary distribution.

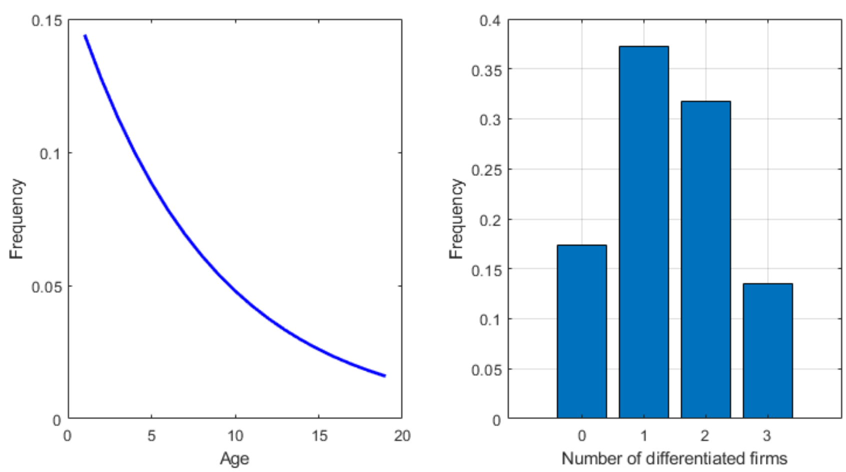
Figure 8 shows the distribution of firm age and the distribution of the number of differentiated firms in the industry. Due to the assumption that firms exit with an exogenous probability κ, the frequency of firms with age a is $1-\kappa$ times the frequency of firms aged $a-1$ (the ones that will survive). The fact that the frequency of industries with three firms is relatively low reassures us that we are not making a significant error by limiting the model to a maximum of three firms.

Figure 7: Customer bases in industries with two firms



Notes. This figure displays the customer bases of the two firms in industries with two firms, with the customer base of the older firm in the horizontal axis, and the customer base of the younger firm in the vertical axis. The size of the dots is proportional to the frequency of the industry state in the stationary distribution.

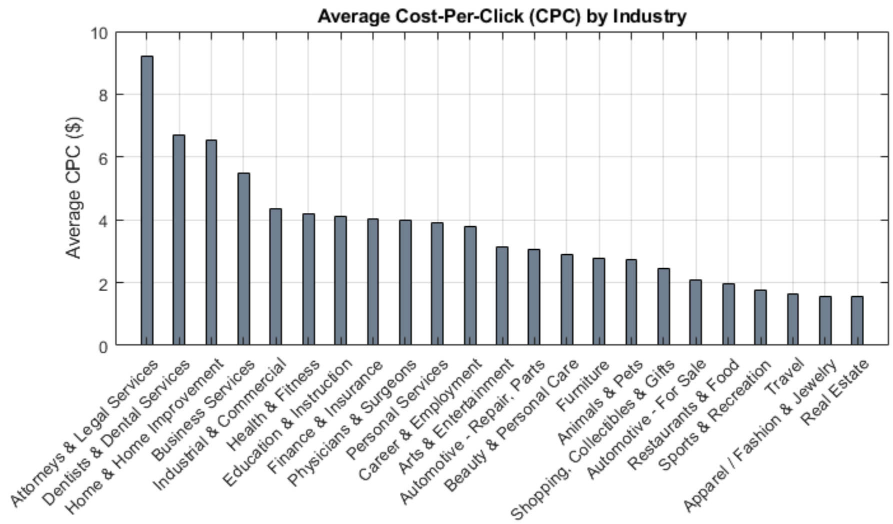
Figure 8: The distributions of firm age and number of firms



Notes. This figure displays the distribution of firm age in the left panel and the distribution of the number of differentiated firms in the right panel. Age is shown up to age 19, as age 20 includes all firms aged 20 or more.

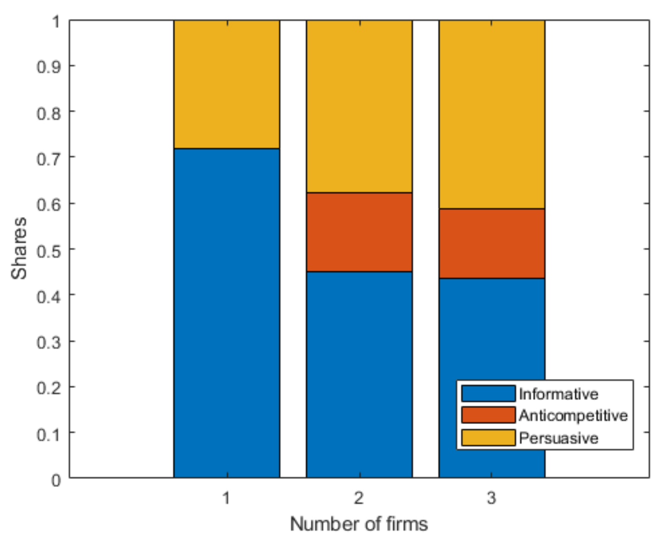
6.3 Additional Figures

Figure 9: Average Coct-Per-Click (CPC) in Google search ads by industry



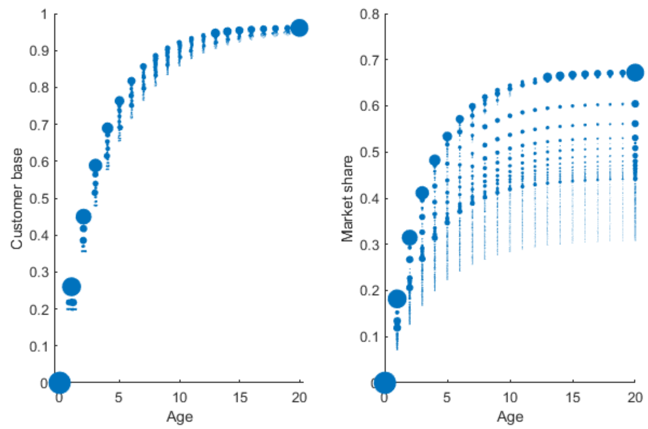
Notes. Adapted from Wordstream (2023). This figure displays the average CPC in Google ads by industry, calculated by dividing the overall cost of a campaign by the number of clicks it received. Each individual click has a different cost as it's determined by the Google Ads auction algorithm.

Figure 10: Decomposition of advertising incentives by motive conditional on age



Notes. This figure displays the average shares of the incentives from the FOC attributable to each motive, weighted by industry advertising expenditure and conditioned on the number of differentiated firms in the industry.

Figure 11: Customer base and market share and age



Notes. This figure displays the relationship of age with customer base (left panel) and market share (right), with the size of the dots indicates the share of this firm type in the stationary distribution.

6.4 Preferences

$$\max_{\{\{c_{\ell,j,i,t}\},a_{\ell,t+1},N_{\ell,t},T_{F,\ell,t},T_{\ell,i,t}\}} U_\ell = \sum_{t=0}^{\infty} \beta^t \left[\frac{C_{\ell t}^{1-\theta}-1}{1-\theta} + L_{\ell t}\right]$$

$$\text{s.t. } C_{\ell,t} = \left(\int_0^1 C_{\ell,i,t}^{\frac{\chi-1}{\chi}} di\right)^{\frac{\chi}{\chi-1}}, \qquad L_{\ell t} = v\left(Q_t T_{\ell,i,t} - \frac{T_{\ell,t}^2}{2}\right)$$

$$C_{\ell,i,t} = \left(c_{\ell,0,i,t}^{\frac{\sigma-1}{\sigma}} + \sum_{j\in\mathcal{I}_{\ell,i,t}} \omega_{\ell,j,i,t} c_{\ell,j,i,t}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

$$w_t N + r_t a_{\ell t} = \int_0^1 \sum_{j\in\mathcal{I}_{\ell,i,t}} c_{\ell,j,i,t} p_{j,i,t} di + a_{\ell,t+1} - a_{\ell,t}$$

(for the case $\theta = 1$, $\lim_{\theta\to 1}\frac{c^{1-\theta}}{1-\theta} = \lim_{\theta\to 1}\frac{c^{1-\theta}-1}{1-\theta} + \lim_{\theta\to 1}\frac{1}{1-\theta} = \ln c + \lim_{\theta\to 1}\frac{1}{1-\theta}$)
We can already plug $C_{\ell,t}$ into the objective function.
The FOCs read:

$$[c_{\ell jt}]: \qquad \frac{\partial U_{\ell,t}}{\partial C_{\ell,t}}\frac{\partial C_{\ell,t}}{\partial C_{\ell,i,t}}\frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = \mu_{\ell,t} p_{j,i,t}$$

where $\frac{\partial U_{\ell,t}}{\partial C_{\ell,t}} = \beta^t C_{\ell,t}^{-\theta}$, $\frac{\partial C_{\ell,t}}{\partial C_{\ell,i,t}} = C_{\ell,t}^{\frac{1}{\chi}} C_{\ell,i,t}^{-\frac{1}{\chi}}$, and $\frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = C_{\ell,i,t}^{\frac{1}{\sigma}} c_{\ell,j,i,t}^{-\frac{1}{\sigma}} \omega_{\ell,j,i,t}$.

We can break down the FOC into three conditions, by defining $P_{\ell,i,t}$ as $P_{\ell,i,t}C_{\ell,i,t} = \sum_{j\in\mathcal{I}_{\ell,i,t}} c_{\ell,j,i,t}p_{j,i,t}$, and $P_{\ell,t}$ as $P_{\ell,t}C_{\ell,t} = \int_0^1 C_{\ell,i,t}P_{i,t}di$:

1. $\frac{\partial U_{\ell,t}}{\partial C_{\ell,t}}\frac{\partial C_{\ell,t}}{\partial c_{\ell,j,i,t}} = \mu_{\ell,t}\frac{\partial E_{\ell,t}}{\partial C_{\ell,t}}\frac{\partial C_{\ell,t}}{\partial c_{\ell,j,i,t}} \implies \left[\frac{\partial U_{\ell,t}}{\partial C_{\ell,t}} - \mu_{\ell,t}\frac{\partial E_{\ell,t}}{\partial C_{\ell,t}}\right]\frac{\partial C_{\ell,t}}{\partial c_{\ell,j,i,t}} = 0 \implies \beta^t C_{\ell,t}^{-\theta} = \mu_{\ell,t}P_{\ell,t}$

2. $\frac{\partial U_{\ell,t}}{\partial C_{\ell,t}}\frac{\partial C_{\ell,t}}{\partial C_{\ell,i,t}}\frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = \mu_{\ell,t}\frac{\partial E_{\ell,t}}{\partial C_{\ell,i,t}}\frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} \implies \left[\frac{\partial U_{\ell,t}}{\partial C_{\ell,t}}\frac{\partial C_{\ell,t}}{\partial C_{\ell,i,t}} - \mu_{\ell,t}\frac{\partial E_{\ell,t}}{\partial C_{\ell,i,t}}\right]\frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = 0$, where using from the previous condition that $\mu_{\ell,t} = \beta^t C_{\ell,t}^{-\theta}P_{\ell,t}^{-1}$, we get: $C_{\ell,t}^{\frac{1}{\chi}}C_{\ell,i,t}^{-\frac{1}{\chi}} = \frac{P_{\ell,i,t}}{P_{\ell,t}} \implies C_{\ell,i,t} = C_{\ell,t}\left(\frac{P_{\ell,t}}{P_{\ell,i,t}}\right)^{\chi}$,

and plugging it into the definition of $C_{\ell,t}$, we get $P_{\ell,t} = \left(\int_0^1 P_{\ell,i,t}^{1-\chi} di\right)^{\frac{1}{1-\chi}}$. Note that if Cobb-Douglas (i.e. $\chi = 1$), then $E_{\ell,t} = P_{\ell,t}C_{\ell,t} = P_{\ell,i,t}C_{\ell,i,t}$.

3. $\frac{\partial U_{\ell,t}}{\partial C_{\ell,t}} \frac{\partial C_{\ell,t}}{\partial C_{\ell,i,t}} \frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = \mu_{\ell,t} \frac{\partial E_{\ell,t}}{\partial c_{\ell,j,i,t}} = \mu_{\ell,t} p_{j,i,t}$, where using from the previous conditions that $\mu_{\ell,t} = \beta^t C_{\ell,t}^{-\theta} P_{\ell,t}^{-1}$ and $C_{\ell,t}^{\frac{1}{\chi}} C_{\ell,i,t}^{-\frac{1}{\chi}} = \frac{P_{\ell,i,t}}{P_{\ell,t}}$, we get: $\frac{P_{\ell,i,t}}{P_{\ell,t}} C_{\ell,i,t}^{\frac{1}{\sigma}} c_{\ell,j,i,t}^{-\frac{1}{\sigma}} \omega_{\ell,j,i,t} = \frac{p_{j,i,t}}{P_{\ell,t}} \implies c_{\ell,j,i,t} = C_{\ell,i,t} \left(\frac{\omega_{\ell,j,i,t} P_{\ell,i,t}}{p_{j,i,t}}\right)^{\sigma}$, and plugging it into the definition of $C_{\ell,i,t}$, we get: $P_{\ell,i,t} = \left(p_{0,i,t}^{1-\sigma} + \sum_{j \in \mathcal{I}_{\ell,i,t}} \omega_{\ell,j,i,t}^{\sigma} p_{j,i,t}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$

The FOC for assets is:

$$[a_{\ell,t+1}]: \qquad \mu_{\ell t} = (1 + r_{t+1})\mu_{\ell,t+1}$$

From the first one, using that $\mu_{\ell,t} = \beta^t C_{\ell,t}^{-\theta} P_{\ell,t}^{-1}$, we get the Euler equation: $\beta^t C_{\ell,t}^{-\theta} P_{\ell,t}^{-1} = (1 + r_{t+1})\beta^{t+1} C_{\ell,t+1}^{-\theta} P_{\ell,t+1}^{-1}$, which assuming $\theta = 1$ (i.e. logarithmic preferences on $C_{\ell,t}$), then the expenditure choice is independent of the price indices (so, the awareness set just affects the intratemporal allocation of expenditure).

So, assuming $\chi = \theta = 1$, we have:

$$c_{\ell,j,i,t} = E_{\ell,t} P_{\ell,i,t}^{\sigma-1} p_{j,i,t}^{-\sigma} \omega_{\ell,j,i,t}^{\sigma}$$
$$\frac{E_{\ell,t+1}}{E_{\ell,t}} = \beta(1 + r_{t+1})$$

where E_ℓ is the expenditure of individual ℓ. So, the growth of expenditure is symmetric for all individuals (and the level is also identical if all individuals start with the same level of assets).

Since the individual is characterised by the awareness set, from now on I use the subindex $\mathcal{I}$, instead of ℓ. The share of expenditure of each consumer on each good they know is: $s_{\mathcal{I},j} = \frac{p_j c_{\mathcal{I},j}}{E} = P_{\mathcal{I}}^{\sigma-1} p_j^{1-\sigma} \omega_j^{\sigma} = p_j^{1-\sigma} \omega_j^{\sigma} \left[p_{0,t}^{1-\sigma} + \sum_{k \in \mathcal{I}_\ell} \omega_{\ell,k}^{\sigma} p_k^{1-\sigma}\right]^{-1} = \left[\left(\frac{p_{0,t}}{p_j}\right)^{1-\sigma} \left(\frac{1}{\omega_j}\right)^{\sigma} + \sum_{k \in \mathcal{I}_\ell} \left(\frac{p_{k,t}}{p_j}\right)^{1-\sigma} \left(\frac{\omega_k}{\omega_j}\right)^{\sigma}\right]^{-1}$
So, using the definition of markup $\mathcal{M}_j = \frac{p_j A_j}{w}$:

$$s_{\mathcal{I},j} = \left[\left(\frac{1}{\mathcal{M}_j A_0}\right)^{1-\sigma} \left(\frac{1}{\omega_j}\right)^{\sigma} + \sum_{k \in \mathcal{I}_\ell} \left(\frac{\mathcal{M}_k A_j}{\mathcal{M}_j A_k}\right)^{1-\sigma} \left(\frac{\omega_k}{\omega_j}\right)^{\sigma}\right]^{-1}$$

Next, the choice of media time is straightforward: $\frac{\partial L_{\ell,t}}{\partial T_{\ell,t}} = \upsilon \left(Q_t - T_{\ell,t}\right)$, so $T_t = Q_t$. And so, optimal leisure as a function of Q is: $L_t^* = \upsilon \frac{Q_t^2}{2}$.

6.5 Proof that $\delta = 0$ allows a simpler state

Setting $\delta = 0$ allows a simpler sufficient state:
If individuals don't die (or forget), then, given that learning is independent for each good,

we have that a sufficient information that allows us to identify the industry state is the mass of consumers that are aware of good j for each $j \in \mathcal{J}_{i,t}$. Intuitively, the reason why $\delta > 0$ doesn't allow this simplification is that the fact that consumers die with positive probability breaks the independence of the events of being aware of a particular good. That is: given that the older the consumer the more likely she is aware of the good, then, knowing that the consumer is aware of a good allows us to get a better guess of the consumer's age.

Proposition 2 *If $\delta = 0$ and $M_{j,it}$ is the mass of consumers aware of good j, then the distribution of consumers over the awareness sets is given by* $M_{i,t}(\mathcal{I}) = \prod\limits_{j \in \mathcal{I}} M_{j,it} \prod\limits_{j \notin \mathcal{I}} (1 - M_{j,it})$

Proof. By induction on t. Set $t = 0$ as the first period a differentiated firm enters. Then, it is trivially satisfied for $t = 0$. We'll see that if it is true for $t - 1$ that $M_{i,t-1}(\mathcal{I}) = \prod\limits_{j \in \mathcal{I}} M_{j,it-1} \prod\limits_{j \notin \mathcal{I}} (1 - M_{j,it-1})$, then it is true for t that $M_{i,t}(\mathcal{I}) = \prod\limits_{j \in \mathcal{I}} M_{j,it} \prod\limits_{j \notin \mathcal{I}} (1 - M_{j,it})$:
By the law of motion of consumers:

$$M_{i,t}(\mathcal{I}) = (1 - \delta) \sum_{\mathcal{I}' \subseteq \mathcal{I}} M_{i,t-1}(\mathcal{I}') \prod_{k \in \mathcal{I} \setminus \mathcal{I}'} \rho_{k,i,t-1} \prod_{k \notin \mathcal{I}} (1 - \rho_{k,i,t-1})$$

Using the induction hypothesis:

$$M_{i,t}(\mathcal{I}) = (1 - \delta) \sum_{\mathcal{I}' \subseteq \mathcal{I}} \prod_{j \in \mathcal{I}'} M_{j,it-1} \prod_{j \notin \mathcal{I}'} (1 - M_{j,it-1}) \prod_{k \in \mathcal{I} \setminus \mathcal{I}'} \rho_{k,i,t-1} \prod_{k \notin \mathcal{I}} (1 - \rho_{k,i,t-1})$$

Note that $\prod\limits_{j \notin \mathcal{I}'} (1 - M_{j,it-1}) = \prod\limits_{j \in \mathcal{I} \setminus \mathcal{I}'} (1 - M_{j,it-1}) \prod\limits_{j \notin \mathcal{I}} (1 - M_{j,it-1})$; so, we can write it as:

$$M_{i,t}(\mathcal{I}) = (1 - \delta) \sum_{\mathcal{I}' \subseteq \mathcal{I}} \prod_{j \in \mathcal{I}'} M_{j,it-1} \prod_{k \in \mathcal{I} \setminus \mathcal{I}'} [(1 - M_{k,it-1}) \rho_{k,i,t-1}] \prod_{k \notin \mathcal{I}} [(1 - M_{k,it-1})(1 - \rho_{k,i,t-1})]$$

And using that it holds $\prod\limits_{j \in \mathcal{J}} (a_j + b_j) = \sum\limits_{\mathcal{I} \subseteq \mathcal{J}} \prod\limits_{j \in \mathcal{I}} a_j \prod\limits_{j \notin \mathcal{I}} b_j$; then, we have:

$$M_{i,t}(\mathcal{I}) = (1 - \delta) \prod_{j \in \mathcal{I}} [M_{j,it-1} + (1 - M_{j,it-1}) \rho_{j,i,t-1}] \prod_{k \notin \mathcal{I}} [(1 - M_{k,it-1})(1 - \rho_{k,i,t-1})]$$

On the other hand, the law of motion of $M_{k,it}$ is:
$M_{k,it} = (1 - \delta) M_{k,it-1} + (1 - \delta)(1 - M_{k,it-1}) \rho_{k,i,t-1} \implies (1 - M_{k,it-1})(1 - \rho_{k,i,t-1}) = 1 - \frac{M_{k,it}}{1 - \delta}$.
So, if $\delta = 0$, we have, as wanted:

$$M_{i,t}(\mathcal{I}) = \prod_{j \in \mathcal{I}} M_{j,it} \prod_{j \notin \mathcal{I}} (1 - M_{j,it})$$

$\blacksquare$

6.6 Production Firms

6.6.1 Derivatives of profits and expenditure shares

1. $\pi_{j,i} = (1-\mathcal{M}_{j,i}^{-1})\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I}) s_{\mathcal{I},j,i}$

   (a) $\frac{\partial \pi_{j,i}}{\partial e_{k,i}} = (1-\mathcal{M}_{j,i}^{-1})\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\partial s_{\mathcal{I},j,i}}{\partial e_{k,i}}$, $k \in \mathcal{J}_i$

   (b) $\frac{\partial \pi_{j,i}}{\partial \mathcal{M}_{j,i}} = \frac{s_{j,i}}{\mathcal{M}_{j,i}^2} + (1-\mathcal{M}_{j,i}^{-1})\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\partial s_{\mathcal{I},j,i}}{\partial \mathcal{M}_{j,i}}$

   (c) $\frac{\partial \pi_{j,i}}{\partial \mathcal{M}_{k,i}} = (1-\mathcal{M}_{j,i}^{-1})\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\partial s_{\mathcal{I},j,i}}{\partial \mathcal{M}_{k,i}}$, $k \neq j$

2. $s_{\mathcal{I},j,i} = \frac{p_{j,i}c_{\mathcal{I},j,i}}{E} = \frac{p_{j,i}^{1-\sigma}\omega_{j,i}^{\sigma}}{P_{\mathcal{I},i}^{1-\sigma}} = \frac{p_{j,i}^{1-\sigma}\omega_{j,i}^{\sigma}}{p_{0,i}^{1-\sigma}+\sum_{k\in\mathcal{I}}\omega_{k,i}^{\sigma}p_{k,i}^{1-\sigma}} = \frac{\mathcal{M}_{j,i}^{1-\sigma}\omega_{j,i}^{\sigma}}{(\frac{\mathcal{M}_{0,i}}{A_0})^{1-\sigma}+\sum\limits_{k\in\mathcal{I}}\omega_{k,i}^{\sigma}\mathcal{M}_{k,i}^{1-\sigma}}$, and $s_{\mathcal{I},j,i}=0$ if $j\notin\mathcal{I}$. So:[19]

   (a) $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{j,i}} = s_{\mathcal{I},j,i}(1-s_{\mathcal{I},j,i})\frac{\sigma}{\omega_{j,i}}$

   (b) $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{k,i}} = -s_{\mathcal{I},j,i}s_{\mathcal{I},k,i}\frac{\sigma}{\omega_{k,i}}$

   (c) $\frac{\partial s_{\mathcal{I},j,i}}{\partial \mathcal{M}_{j,i}} = -s_{\mathcal{I},j,i}(1-s_{\mathcal{I},j,i})\frac{\sigma-1}{\mathcal{M}_{j,i}} = -\frac{\sigma}{\sigma-1}\frac{\mathcal{M}_{j,i}}{\omega_{j,i}}\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{j,i}}$

   (d) $\frac{\partial s_{\mathcal{I},j,i}}{\partial \mathcal{M}_{k,i}} = s_{\mathcal{I},j,i}s_{\mathcal{I},k,i}\frac{\sigma-1}{\mathcal{M}_{k,i}} = -\frac{\sigma}{\sigma-1}\frac{\mathcal{M}_{k,i}}{\omega_{k,i}}\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{k,i}}$

3. $\omega_{j,i} = 1+\nu_s(\alpha_{j,i}T)^{\nu_c}$, $\alpha_{j,i} = \frac{e_{j,i}}{p_{a,i}}$, where $p_{a,i} = \frac{\sum\limits_k e_{k,i}}{\alpha_i}$ if limited ad space is binding, otherwise $p_{a,i} = \bar{p}_{a,i}$.

   (a) $\frac{\partial \omega_{j,i}}{\partial e_{j,i}} = \nu_c\nu_s T^{\nu_c}\alpha_{j,i}^{\nu_c-1}\frac{\partial \alpha_{j,i}}{\partial e_{j,i}}$

   (b) $\frac{\partial \omega_{k,i}}{\partial e_{j,i}} = \nu_c\nu_s T^{\nu_c}\alpha_{k,i}^{\nu_c-1}\frac{\partial \alpha_{k,i}}{\partial e_{j,i}}$

   (c) $\frac{\partial \alpha_{k,i}}{\partial e_{j,i}} = -\frac{\alpha_{k,i}}{\sum\limits_k e_{k,i}}$ if the limited ad space is binding, otherwise $\frac{\partial \alpha_{k,i}}{\partial e_{j,i}} = 0$.

   (d) $\frac{\partial \alpha_{j,i}}{\partial e_{j,i}} = \frac{\alpha_i}{\sum\limits_k e_{k,i}} - \frac{\alpha_{j,i}}{\sum\limits_k e_{k,i}} = \sum\limits_{j'\neq j}\frac{\alpha_{j',i}}{\sum\limits_k e_{k,i}}$ if the limited ad space is binding, otherwise $\frac{\partial \alpha_{j,i}}{\partial e_{j,i}} = \frac{1}{p_{a,i}}$.

So: $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1})\sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\sum\limits_{k\in\mathcal{I}}\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{k,i}}\frac{\partial \omega_{k,i}}{\partial e_{j,i}}$. And using the above expressions we have:

- for $k \neq j$: $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{k,i}}\frac{\partial \omega_{k,i}}{\partial e_{j,i}} = -s_{\mathcal{I},j,i}s_{\mathcal{I},k,i}\frac{\sigma}{\omega_{k,i}}\nu_c\nu_s T^{\nu_c}\alpha_{k,i}^{\nu_c-1}\frac{\partial \alpha_{k,i}}{\partial e_{j,i}}$.
- $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{j,i}}\frac{\partial \omega_{j,i}}{\partial e_{j,i}} = s_{\mathcal{I},j,i}(1-s_{\mathcal{I},j,i})\frac{\sigma}{\omega_{j,i}}\nu_c\nu_s T^{\nu_c}\alpha_{j,i}^{\nu_c-1}\frac{\partial \alpha_{j,i}}{\partial e_{j,i}}$.

[19] Also, in terms of relative consumption: $s_{\mathcal{I},j,i} = \frac{p_{j,i}c_{\mathcal{I},j,i}}{E} = \frac{c_{\mathcal{I},j,i}^{\frac{\sigma-1}{\sigma}}\omega_j P_{\mathcal{I},j,i}^{\frac{\sigma-1}{\sigma}}}{E_{\mathcal{I},j,i}^{\frac{\sigma-1}{\sigma}}} = \frac{c_{\mathcal{I},j,i}^{\frac{\sigma-1}{\sigma}}\omega_j}{Y_{\mathcal{I},j,i}^{\frac{\sigma-1}{\sigma}}}$

So: $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1}) \sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\sigma\nu_s\nu_c T^{\nu_c} s_{\mathcal{I},j,i}\left[\sum\limits_{k\in\mathcal{I}\setminus\{j\}} \frac{s_{\mathcal{I},k,i}\alpha_{k,i}^{\nu_c-1}}{\omega_{k,i}}\left(-\frac{\partial\alpha_{k,i}}{\partial e_{j,i}}\right) + \frac{(1-s_{\mathcal{I},j,i})\alpha_{j,i}^{\nu_c-1}}{\omega_{j,i}}\frac{\partial \alpha_{j,i}}{\partial e_{j,i}}\right]$

And substituting $\frac{\partial \alpha_{k,i}}{\partial e_{j,i}}$:

1. With binding limited ad space (i.e. $J>1$):
   $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1}) \sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\sigma\nu_s\nu_c}{\sum\limits_k e_{k,i}} T^{\nu_c} s_{\mathcal{I},j,i}\left[\sum\limits_{k\in\mathcal{I}\setminus\{j\}} \frac{s_{\mathcal{I},k,i}\alpha_{k,i}^{\nu_c}}{\omega_{k,i}} + \frac{(1-s_{\mathcal{I},j,i})\alpha_{j,i}^{\nu_c-1}}{\omega_{j,i}}(\alpha_i-\alpha_{j,i})\right]$
   Equivalently:
   $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1}) \sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\sigma\nu_s\nu_c}{\sum\limits_k e_{k,i}} T^{\nu_c} s_{\mathcal{I},j,i}\left[\sum\limits_{k\in\mathcal{I}} \frac{s_{\mathcal{I},k,i}\alpha_{k,i}^{\nu_c}}{\omega_{k,i}} + \frac{\alpha_{j,i}^{\nu_c-1}}{\omega_{j,i}}(\alpha_i(1-s_{\mathcal{I},j,i})-\alpha_{j,i})\right]$
   Or equivalently:
   $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1}) \sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\sigma\nu_c}{\sum\limits_k e_{k,i}} s_{\mathcal{I},j,i}\left[\sum\limits_{k\in\mathcal{I}} \frac{s_{\mathcal{I},k,i}\hat{\omega}_{k,i}}{\omega_{k,i}} + \frac{\hat{\omega}_{j,i}}{\omega_{j,i}}\left(\alpha_i\frac{1-s_{\mathcal{I},j,i}}{\alpha_{j,i}}-1\right)\right]$, where $\hat{\omega}_{j,i} = \nu_s(T\alpha_{j,i})^{\nu_c}$

2. With non-binding ad space (i.e. $J=1$):
   $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1}) \sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\sigma\nu_s\nu_c T^{\nu_c} s_{\mathcal{I},j,i}\frac{(1-s_{\mathcal{I},j,i})\alpha_{j,i}^{\nu_c-1}}{\omega_{j,i}p_{a,i}}$

6.6.2 Extension: Persuasive effect as in Cavenaile et al.

Here, $\omega_{\mathcal{I},j,i,t} = \frac{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}{\frac{1}{\#\mathcal{I}}\sum\limits_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}$. With this, now the derivatives with respect to $e_{j,i,t}$ write:

$$\frac{\partial \omega_{\mathcal{I},j,i}}{\partial e_{j,i}} = \omega_{\mathcal{I},j,i}\frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}\frac{\nu_c}{e_{j,i}} - \omega_{\mathcal{I},j,i}\frac{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}{\sum\limits_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}\frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}\frac{\nu_c}{e_{j,i}} \tag{30}$$

$$= \omega_{\mathcal{I},j,i}\frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}\frac{\nu_c}{e_{j,i}}\frac{\sum\limits_{j'\neq j}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}{\sum\limits_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})} \tag{31}$$

$$\frac{\partial \omega_{\mathcal{I},k,i}}{\partial e_{j,i}} = -\omega_{\mathcal{I},k,i}\frac{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}{\sum\limits_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}\frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}\frac{\nu_c}{e_{j,i}} \tag{32}$$

And recall that $\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1})\sum\limits_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{\partial s_{\mathcal{I},j,i}}{\partial e_{j,i}}$, $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{\mathcal{I},j,i}} = s_{\mathcal{I},j,i}(1-s_{\mathcal{I},j,i})\frac{\sigma}{\omega_{\mathcal{I},j,i}}$, and $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{\mathcal{I},k,i}} = -s_{\mathcal{I},j,i}s_{\mathcal{I},k,i}\frac{\sigma}{\omega_{\mathcal{I},k,i}}$; so:

$$\frac{\partial \pi_{j,i}}{\partial e_{j,i}} = (1-\mathcal{M}_{j,i}^{-1})\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\sum_{k\in\mathcal{I}}\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{\mathcal{I},k,i}}\frac{\partial \omega_{\mathcal{I},k,i}}{\partial e_{j,i}}$$

And using the previous expressions we have:

- $\frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{\mathcal{I},j,i}}\frac{\partial \omega_{\mathcal{I},j,i}}{\partial e_{j,i}} = \sigma s_{\mathcal{I},j,i}(1-s_{\mathcal{I},j,i})\frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}\frac{\nu_c}{e_{j,i}}\frac{\sum\limits_{j'\neq j}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}{\sum\limits_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}$

- $\sum_{k\neq j} \frac{\partial s_{\mathcal{I},j,i}}{\partial \omega_{\mathcal{I},k,i}} \frac{\partial \omega_{\mathcal{I},k,i}}{\partial e_{j,i}} = \sum_{k\neq j} \sigma s_{\mathcal{I},j,i} s_{\mathcal{I},k,i} \frac{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}{\sum_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})} \frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}} \frac{\nu_c}{e_{j,i,t}}$
  $= \sigma s_{\mathcal{I},j,i}(1 - s_{\mathcal{I},j,i} - s_{\mathcal{I},0,i}) \frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}} \frac{\nu_c}{e_{j,i,t}} \frac{1+\nu_s(T_t e_{j,i,t})^{\nu_c}}{\sum_{j'\in\mathcal{I}}(1+\nu_s(T_t e_{j',i,t})^{\nu_c})}$

So, adding them, we get that:

$$\frac{\partial \pi_{j,i,t}}{\partial e_{j,i,t}} = (1 - \mathcal{M}_{j,i,t}^{-1}) \sum_{\mathcal{I}\in\mathcal{P}_{j,i,t}} M_{i,t}(\mathcal{I}) \sigma s_{\mathcal{I},j,i} \frac{\nu_s(T_t e_{j,i,t})^{\nu_c}}{1+\nu_s(T_t e_{j,i,t})^{\nu_c}} \frac{\nu_c}{e_{j,i}} \left(1 - s_{\mathcal{I},j,i,t} - s_{\mathcal{I},0,i,t}\frac{\omega_{\mathcal{I},j,i,t}}{\#\mathcal{I}}\right) \tag{33}$$

6.6.3 The effect of learning about another good on $s_{\mathcal{I},j,i}$ and the demand elasticity

Proposition 3 *If $j \in \mathcal{I} \subset \mathcal{I}'$, then:*

1. $s_{\mathcal{I},j,i} > s_{\mathcal{I}',j,i}$
2. $|\epsilon_{\mathcal{I},j,i}| < |\epsilon_{\mathcal{I}',j,i}|$, *where* $\epsilon_{\mathcal{I},j,i} = \frac{p_{j,i}}{c_{\mathcal{I},j,i}} \frac{\partial c_{\mathcal{I},j,i}}{\partial p_{j,i}}$

Proof. If $j \in \mathcal{I} \subset \mathcal{I}'$, then, since $\omega_{k,i}, \mathcal{M}_{k,i} > 0$ for all firm k and $\sigma > 1$:

$$s_{\mathcal{I},j,i} = \left[(A_0\mathcal{M}_{j,i})^{\sigma-1}\omega_{j,i}^{-\sigma} + \sum_{k\in\mathcal{I}} \left(\frac{\omega_{k,i}}{\omega_{j,i}}\right)^{\sigma} \left(\frac{\mathcal{M}_{j,i}}{\mathcal{M}_{k,i}}\right)^{\sigma-1}\right]^{-1}$$
$$> \left[(A_0\mathcal{M}_{j,i})^{\sigma-1}\omega_{j,i}^{-\sigma} + \sum_{k\in\mathcal{I}} \left(\frac{\omega_{k,i}}{\omega_{j,i}}\right)^{\sigma} \left(\frac{\mathcal{M}_{j,i}}{\mathcal{M}_{k,i}}\right)^{\sigma-1} + \sum_{k\in\mathcal{I}'\setminus\mathcal{I}} \left(\frac{\omega_{k,i}}{\omega_{j,i}}\right)^{\sigma} \left(\frac{\mathcal{M}_{j,i}}{\mathcal{M}_{k,i}}\right)^{\sigma-1}\right]^{-1} = s_{\mathcal{I}',j,i}$$

For 2, define $\epsilon_{\mathcal{I},j,i} = \frac{p_{j,i}}{c_{\mathcal{I},j,i}} \frac{\partial c_{\mathcal{I},j,i}}{\partial p_{j,i}}$, and note that $\frac{\partial s_{\mathcal{I},j,i}}{\partial \mathcal{M}_{j,i}} = \frac{\partial s_{\mathcal{I},j,i}}{\partial p_{j,i}} \frac{\partial p_{j,i}}{\partial \mathcal{M}_{j,i}} = \frac{1}{E}\left[c_{\mathcal{I},j,i} + p_{j,i}\frac{\partial c_{\mathcal{I},j,i}}{\partial p_{j,i}}\right] w = \frac{s_{\mathcal{I},j,i}}{\mathcal{M}_{j,i}}(1+\epsilon_{\mathcal{I},j,i})$. And using the expression for $\frac{\partial s_{\mathcal{I},j,i}}{\partial \mathcal{M}_{j,i}}$ in 6.6.1, we have:

$$-\frac{s_{\mathcal{I},j,i}}{\mathcal{M}_{j,i}}(\sigma-1)(1-s_{\mathcal{I},j,i}) = \frac{s_{\mathcal{I},j,i}}{\mathcal{M}_{j,i}}(1+\epsilon_{\mathcal{I},j,i}) \implies \epsilon_{\mathcal{I},j,i} = -\sigma + s_{\mathcal{I},j,i}(\sigma-1)$$

So, using 1, we have $j \in \mathcal{I} \subset \mathcal{I}'$ $0 > \epsilon_{\mathcal{I},j,i} > \epsilon_{\mathcal{I}',j,i}$ ■

6.6.4 Derivation of optimal markup:

$0 = \frac{\partial \pi_{j,i}}{\partial p_{j,i}} = y_{j,i} + \left(p_{j,i} - \frac{\partial w N_{j,i}}{\partial N_{j,i}}\frac{\partial N_{j,i}}{\partial y_{j,i}}\right) \sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I}) \frac{\partial c_{\mathcal{I},j,i}}{\partial p_{j,i}}$

On the one hand: $\frac{\partial c_{\mathcal{I},j,i}}{\partial p_{j,i}} = c_{\mathcal{I},j,i}\left[-\sigma p_{j,i}^{-1} + (\sigma-1)P_{\mathcal{I},i}^{-1}\frac{\partial P_{\mathcal{I},i}}{\partial p_{j,i}}\right] = \frac{c_{\mathcal{I},j,i}}{p_{j,i}}\left[-\sigma + (\sigma-1)\frac{p_{j,i}}{P_{\mathcal{I},i}}\frac{\partial P_{\mathcal{I},i}}{\partial p_{j,i}}\right]$.

And $\frac{\partial P_{\mathcal{I},i}}{\partial p_{j,i}} = P_{\mathcal{I},i}^{\sigma} p_{j,i}^{-\sigma}\omega_{j,i}^{\sigma}$. So, we have:

$0 = y_{j,i} + (p_{j,i} - w)\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I})\frac{c_{\mathcal{I},j,i}}{p_{j,i}}\left[-\sigma + (\sigma-1)\left(\frac{P_{\mathcal{I},i}}{p_{j,i}}\right)^{\sigma-1}\omega_{j,i}^{\sigma}\right]$

Since $s_{\mathcal{I},j,i} = \left(\frac{P_{\mathcal{I},i}}{p_{j,i}}\right)^{\sigma-1}\omega_{j,i}^{\sigma}$ and multiplying by $\frac{p_{j,i}}{E}$:

$0 = s_{j,i} + \left(1 - \mathcal{M}_{j,i}^{-1}\right)\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} M_i(\mathcal{I}) s_{\mathcal{I},j,i}\left[-\sigma + (\sigma-1)s_{\mathcal{I},j,i}\right]$

Equivalently, we can write it:
$0 = 1 + \left(1 - \frac{1}{\mathcal{M}_{j,i}}\right) \sum_{\mathcal{I}\in\mathcal{P}_{j,i}} \frac{M_i(\mathcal{I})c_{\mathcal{I},j,i}}{y_{j,i}} \left[-\sigma + (\sigma-1)s_{\mathcal{I},j,i}\right]$
And using that $\sum_{\mathcal{I}\in\mathcal{P}_{j,i}} \frac{M_i(\mathcal{I})c_{\mathcal{I},j,i}}{y_{j,i}} = 1$ and defining $\bar{s}_{j,i} = \sum_{\mathcal{I}\in\mathcal{P}_{j,i}} \frac{M_i(\mathcal{I})c_{\mathcal{I},j,i}}{y_{j,i}} s_{\mathcal{I},j,i}$:
$1 = \left(1 - \frac{1}{\mathcal{M}_{j,i}}\right) [\sigma - (\sigma-1)\bar{s}_{j,i}] \implies [\sigma - (\sigma-1)\bar{s}_{j,i}]^{-1} = 1 - \frac{1}{\mathcal{M}_{j,i}}$
Rearranging:
$\frac{1}{\mathcal{M}_{j,i}} = \frac{\sigma-1-(\sigma-1)\bar{s}_{j,i}}{\sigma-(\sigma-1)\bar{s}_{j,i}}$

6.7 Derivation of the expression for the FOC:

Derivative of $\rho_{j,i}$ with respect to $e_{k,i}$. $\frac{\partial \rho_k}{\partial e_j} = \psi_c \psi_s T^{\psi_c} \alpha_k^{\psi_c - 1} \frac{\partial \alpha_k}{\partial e_j}$

- If $J = 1$ (ad space not binding), then $\frac{\partial \alpha_k}{\partial e_j} = 0$, so $\frac{\partial \rho_k}{\partial e_j} = 0$; and $\frac{\partial \alpha_j}{\partial e_j} = \frac{1}{p_a}$, so: $\frac{\partial \rho_j}{\partial e_j} = \psi_c \psi_s T^{\psi_c} \alpha_j^{\psi_c-1} \frac{1}{p_a}$
- If $J > 1$ (ad space binding), then $\frac{\partial \alpha_k}{\partial e_j} = -\frac{a_k}{\sum_s e_s}$, so $\frac{\partial \rho_k}{\partial e_j} = -\psi_c \psi_s T^{\psi_c} \alpha_k^{\psi_c-1} \frac{a_k}{\sum_s e_s}$; and $\frac{\partial \alpha_j}{\partial e_j} = \frac{\alpha}{\sum_s e_s} - \frac{\alpha_j}{\sum_s e_s} = \frac{\sum_{k\neq j} \alpha_k}{\sum_s e_s}$, so: $\frac{\partial \rho_j}{\partial e_j} = \psi_c \psi_s T^{\psi_c} \alpha_j^{\psi_c-1} \frac{\sum_{k\neq j}\alpha_k}{\sum_s e_s}$.
  Note that since $\alpha_j = \alpha \frac{e_j}{\sum_k e_k}$, we can rewrite them as: $\frac{\partial \rho_k}{\partial e_j} = -\psi_c \psi_s T^{\psi_c} \frac{\alpha^{\psi_c}}{(\sum_s e_s)^{\psi_c+1}} e_k^{\psi_c-1} e_k$ and $\frac{\partial \rho_j}{\partial e_j} = \psi_c \psi_s T^{\psi_c} \frac{\alpha^{\psi_c}}{(\sum_s e_s)^{\psi_c+1}} e_j^{\psi_c-1} \sum_{k\neq j} e_k$

Derivative of a function of next period vector of masses.

Lemma 1 *If $f : \vec{\hat{M}}' \to \mathbb{R}$, then we have:*

1. $\frac{\partial f}{\partial \rho_j} = \sum\limits_{\mathcal{I}\in\mathcal{P}_{-j}} M(\mathcal{I}) \sum\limits_{\mathcal{I}'\in\mathcal{P}_{-j}, \mathcal{I}'\supseteq\mathcal{I}} \frac{\Theta(\mathcal{I}\to\mathcal{I}')}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}'\cup\{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')}\right]$
2. *For the anticompettive motive, it will be useful:*
$$\frac{\partial f}{\partial \rho_j} = \sum_{\mathcal{I}\in\mathcal{P}_{-k,-j}} M(\mathcal{I}) \sum_{\substack{\mathcal{I}'\in\mathcal{P}_{-k,-j}\\ \mathcal{I}'\supseteq\mathcal{I}}} \frac{\Theta(\mathcal{I}\to\mathcal{I}')}{1-\rho_j} \left[\left[\frac{\partial f}{\partial M'(\mathcal{I}'\cup\{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')}\right] - \left[\frac{\partial f}{\partial M'(\mathcal{I}'\cup\{k,j\})} - \frac{\partial f}{\partial M'(\mathcal{I}'\cup\{k\})}\right]\right]$$
$$+ \sum_{\mathcal{I}\in\mathcal{P}_{-k,-j}} (M(\mathcal{I}) + M(\mathcal{I}\cup\{k\})) \sum_{\substack{\mathcal{I}'\in\mathcal{P}_{-k,-j}\\ \mathcal{I}'\supseteq\mathcal{I}}} \frac{\Theta(\mathcal{I}\cup\{k\}\to\mathcal{I}'\cup\{k\})}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}'\cup\{k,j\})} - \frac{\partial f}{\partial M'(\mathcal{I}'\cup\{k\})}\right]$$

Proof. First, recall that $\hat{M}'(\mathcal{I}') = \sum_{\mathcal{I}\in\mathcal{P}(\mathcal{J})} M(\mathcal{I})\Theta(\mathcal{I}\to\mathcal{I}')$.
Next, the derivatives of $\Theta(\mathcal{I}\to\mathcal{I}')$ wrt ρ_j are:

1. If $\mathcal{I} \nsubseteq \mathcal{I}'$: $\frac{\partial\Theta(\mathcal{I}\to\mathcal{I}')}{\partial\rho_j} = 0$
2. If $\mathcal{I} \subseteq \mathcal{I}'$:
   (a) If $j\in\mathcal{I}$: $\frac{\partial\Theta(\mathcal{I}\to\mathcal{I}')}{\partial\rho_j} = 0$
   (b) If $j\in\mathcal{I}'\setminus\mathcal{I}$: $\frac{\partial\Theta(\mathcal{I}\to\mathcal{I}')}{\partial\rho_j} = (1-\delta)\prod\limits_{k\in\mathcal{I}'\setminus(\mathcal{I}\cup\{j\})} \rho_k \prod\limits_{k\notin\mathcal{I}'} (1-\rho_k) = \frac{\Theta(\mathcal{I}\to\mathcal{I}'\setminus\{j\})}{1-\rho_j}$

(c) If $j \notin \mathcal{I}'$: $\frac{\partial \Theta(\mathcal{I} \to \mathcal{I}')}{\partial \rho_j} = -(1-\delta) \prod_{k \in \mathcal{I}' \setminus \mathcal{I}} \rho_k \prod_{k \notin (\mathcal{I}' \cup \{j\})} (1-\rho_k) = -\frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j}$

Using this, the derivative of $\hat{M}'(\mathcal{I}')$ wrt ρ_j is:

1. If $j \in \mathcal{I}'$: $\frac{\partial \hat{M}'(\mathcal{I}')}{\partial \rho_j} = \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subset \mathcal{I}'} M(\mathcal{I}) \frac{\partial \Theta(\mathcal{I} \to \mathcal{I}')}{\partial \rho_j} = \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subset \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}' \setminus \{j\})}{1-\rho_j}$

2. If $j \notin \mathcal{I}'$: $\frac{\partial \hat{M}'(\mathcal{I}')}{\partial \rho_j} = \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subseteq \mathcal{I}'} M(\mathcal{I}) \frac{\partial \Theta(\mathcal{I} \to \mathcal{I}')}{\partial \rho_j} = -\sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subseteq \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j}$

And the derivative of a generic function $f : \vec{\hat{M}}' \to \mathbb{R}$ wrt ρ_j is:

$$
\begin{aligned}
\frac{\partial f}{\partial \rho_j} &= \sum_{\mathcal{I}' \in \mathcal{P}} \frac{\partial f}{\partial M'(\mathcal{I}')} \frac{\partial M'(\mathcal{I}')}{\partial \rho_j} = \sum_{\mathcal{I}' \in \mathcal{P}_j} \frac{\partial f}{\partial M'(\mathcal{I}')} \frac{\partial M'(\mathcal{I}')}{\partial \rho_j} + \sum_{\mathcal{I}' \in \mathcal{P}_{-j}} \frac{\partial f}{\partial M'(\mathcal{I}')} \frac{\partial M'(\mathcal{I}')}{\partial \rho_j} \\
&= \sum_{\mathcal{I}' \in \mathcal{P}_j} \frac{\partial f}{\partial M'(\mathcal{I}')} \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subset \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}' \setminus \{j\})}{1-\rho_j} - \sum_{\mathcal{I}' \in \mathcal{P}_{-j}} \frac{\partial f}{\partial M'(\mathcal{I}')} \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subseteq \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j}
\end{aligned}
$$

Now we are going to merge the two summations using that $\mathcal{P}_j = \{\mathcal{I} \cup \{j\} | \mathcal{I} \in \mathcal{P}_{-j}\}$ [**Proof:** from any set $\mathcal{I}$ that doesn't contain j we can build one by adding j to $\mathcal{I}$ (that is, $\{\mathcal{I} \cup \{j\} | \mathcal{I} \in \mathcal{P}_{-j}\} \subseteq \mathcal{P}_j$), and that from any $\mathcal{I}'$ that contains j we can build another one that doesn't contain j by removing j from $\mathcal{I}'$ (that is, $\mathcal{P}_j = \{(\mathcal{I}' \setminus \{j\}) \cup \{j\} | \mathcal{I}' \in \mathcal{P}_j\} \subseteq \{\mathcal{I} \cup \{j\} | \mathcal{I} \in \mathcal{P}_{-j}\}$].

Using this in the previous expression, we get:

$$
\begin{aligned}
\frac{\partial f}{\partial \rho_j} &= \sum_{\mathcal{I}' \in \mathcal{P}_{-j}} \frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subseteq \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} - \sum_{\mathcal{I}' \in \mathcal{P}_{-j}} \frac{\partial f}{\partial M'(\mathcal{I}')} \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subseteq \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} \\
&= \sum_{\mathcal{I}' \in \mathcal{P}_{-j}} \sum_{\mathcal{I} \in \mathcal{P}_{-j}, \mathcal{I} \subseteq \mathcal{I}'} M(\mathcal{I}) \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')} \right] \\
&= \sum_{\mathcal{I} \in \mathcal{P}_{-j}} M(\mathcal{I}) \sum_{\mathcal{I}' \in \mathcal{P}_{-j}, \mathcal{I}' \supseteq \mathcal{I}} \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')} \right]
\end{aligned}
$$

where for the last equality, I have used that $\{(\mathcal{I}, \mathcal{I}') | \mathcal{I} \in \mathcal{P}, \mathcal{I}' \subseteq \mathcal{I}\} = \{(\mathcal{I}, \mathcal{I}') | \mathcal{I}' \in \mathcal{P}, \mathcal{I} \supseteq \mathcal{I}'\}$.

This proves the first expression of the lemma. For the second:
First, note that above we have shown that $\{\mathcal{I} \cup \{j\} | \mathcal{I} \in \mathcal{P}_{-j}\} = \mathcal{P}_j$, which implies that $\mathcal{P} = \mathcal{P}_{-j} \cup \{\mathcal{I} \cup \{j\} | \mathcal{I} \in \mathcal{P}_{-j}\}$. Analogously, defining $\mathcal{P}_{-k,-j} = \{\mathcal{I} \in \mathcal{P} | j, k \notin \mathcal{I}\}$, we have: $\mathcal{P}_{-j} = \mathcal{P}_{-k,-j} \cup \{\mathcal{I} \cup \{k\} | \mathcal{I} \in \mathcal{P}_{-k,-j}\}$, so the previous expression becomes

$$
\begin{aligned}
\frac{\partial f}{\partial \rho_j} = \sum_{\mathcal{I} \in \mathcal{P}_{-k,-j}} M(\mathcal{I}) & \sum_{\substack{\mathcal{I}' \in \mathcal{P}_{-j} \\ \mathcal{I}' \supseteq \mathcal{I}}} \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')} \right] \\
+ M(\mathcal{I} \cup \{k\}) & \sum_{\substack{\mathcal{I}' \in \mathcal{P}_{-j} \\ \mathcal{I}' \supseteq \mathcal{I} \cup \{k\}}} \frac{\Theta(\mathcal{I} \cup \{k\} \to \mathcal{I}')}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')} \right]
\end{aligned}
$$

Now, for the first line, I use the following equivalence: for each $\mathcal{I} \in \mathcal{P}_{-k,-j}$ we have $\{\mathcal{I}' \in \mathcal{P}_{-j} | \mathcal{I}' \supseteq \mathcal{I}\} = \{\mathcal{I}' \in \mathcal{P}_{-k,-j} | \mathcal{I}' \supseteq \mathcal{I}\} \cup \{\mathcal{I}' \cup \{k\} | \mathcal{I}' \in \mathcal{P}_{-k,-j}, \mathcal{I}' \supseteq \mathcal{I}\}$. And for the second line, I use the equivalence: for $\mathcal{I} \cup \{k\}$ we have $\{\mathcal{I}' \in \mathcal{P}_{-j} | \mathcal{I}' \supseteq \mathcal{I} \cup \{k\}\} = \{\mathcal{I}' \cup \{k\} | \mathcal{I}' \in \mathcal{P}_{-k,-j}, \mathcal{I}' \supseteq \mathcal{I}\}$.

$$\begin{aligned}\frac{\partial f}{\partial \rho_j} = &\sum_{\mathcal{I} \in \mathcal{P}_{-k,-j}} M(\mathcal{I}) \sum_{\substack{\mathcal{I}' \in \mathcal{P}_{-k,-j} \\ \mathcal{I}' \supseteq \mathcal{I}}} \left[\frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')} \right] + \frac{\Theta(\mathcal{I} \to \mathcal{I}' \cup \{k\})}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k,j\})} - \frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k\})} \right] \right] \\ &+ M(\mathcal{I} \cup \{k\}) \sum_{\substack{\mathcal{I}' \in \mathcal{P}_{-k,-j} \\ \mathcal{I}' \supseteq \mathcal{I}}} \frac{\Theta(\mathcal{I} \cup \{k\} \to \mathcal{I}' \cup \{k\})}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k,j\})} - \frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k\})} \right]\end{aligned}$$

Finally, I use that for $\mathcal{I}, \mathcal{I}' \in \mathcal{P}_{-k,-j}$ with $\mathcal{I}' \supseteq \mathcal{I}$, we have $\Theta(\mathcal{I} \cup \{k\} \to \mathcal{I}' \cup \{k\}) = \prod_{h \in \mathcal{I}' \setminus \mathcal{I}} \rho_h \prod_{h \notin \mathcal{I}' \cup \{k\}} (1-\rho_h) \cdot (\rho_k + 1 - \rho_k) = \Theta(\mathcal{I} \to \mathcal{I}' \cup \{k\}) + \Theta(\mathcal{I} \to \mathcal{I}')$. So, I substitute in the first line $\Theta(\mathcal{I} \to \mathcal{I}' \cup \{k\}) = \Theta(\mathcal{I} \cup \{k\} \to \mathcal{I}' \cup \{k\}) - \Theta(\mathcal{I} \to \mathcal{I}')$.

$$\begin{aligned}\frac{\partial f}{\partial \rho_j} = &\sum_{\mathcal{I} \in \mathcal{P}_{-k,-j}} M(\mathcal{I}) \sum_{\substack{\mathcal{I}' \in \mathcal{P}_{-k,-j} \\ \mathcal{I}' \supseteq \mathcal{I}}} \frac{\Theta(\mathcal{I} \to \mathcal{I}')}{1-\rho_j} \left[\left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{j\})} - \frac{\partial f}{\partial M'(\mathcal{I}')} \right] - \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k,j\})} - \frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k\})} \right] \right] \\ &+ (M(\mathcal{I}) + M(\mathcal{I} \cup \{k\})) \sum_{\substack{\mathcal{I}' \in \mathcal{P}_{-k,-j} \\ \mathcal{I}' \supseteq \mathcal{I}}} \frac{\Theta(\mathcal{I} \cup \{k\} \to \mathcal{I}' \cup \{k\})}{1-\rho_j} \left[\frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k,j\})} - \frac{\partial f}{\partial M'(\mathcal{I}' \cup \{k\})} \right]\end{aligned}$$

■

With uncertainty:

$$\text{For } \mathcal{I}' \in \mathcal{P}(\mathcal{J}'), \qquad M'(\mathcal{I}') = \begin{cases} \sum\limits_{\{\mathcal{I} \in \mathcal{P}(\mathcal{J}) : \mathcal{I} \cap \mathcal{J}' = \mathcal{I}'\}} \hat{M}(\mathcal{I}) & \text{, if } \mathcal{I}' \subseteq \mathcal{J} \\ 0 & \text{, if } \mathcal{I}' \nsubseteq \mathcal{J} \end{cases} \tag{34}$$

where the first case says that two consumers become identical in industry i if all the firms in which they differed exit, whereas the second case says that there are no consumers who are aware of a newborn firm. The last piece of infornation needed to compute expected values is the probabilities that the set of differentiated goods moves from $\mathcal{J}$ to $\mathcal{J}' \subseteq \mathcal{J} \cup \{e\}$, where e denotes an entrant. These probabilities are given by:

$$\text{For } \mathcal{J}' \in \mathcal{P}(\mathcal{J} \cup e)), \qquad Prob\{\mathcal{J} \to \mathcal{J}'\} = \begin{cases} (1 - z_{e,i,t}) \prod\limits_{j \in \mathcal{J} \cap \mathcal{J}'} (1-\kappa) \prod\limits_{j \in \mathcal{J} \setminus \mathcal{J}'} \kappa & \text{, if } e \notin \mathcal{J}' \\ z_{e,i,t} \prod\limits_{j \in \mathcal{J} \cap \mathcal{J}'} (1-\kappa) \prod\limits_{j \in \mathcal{J} \setminus \mathcal{J}'} \kappa & \text{, if } e \in \mathcal{J}' \end{cases} \tag{35}$$

In the model, there is uncertainty on $\mathcal{J}'$, so I am more interested in finding the derivative of the expected value of a function $g : \vec{M}' \to \mathbb{R}$ rather than the derivative of a function $f : \vec{\hat{M}}' \to \mathbb{R}$ (note that f is defined on $\vec{\hat{M}}'$, that is, the next period distribution if there weren't entry and exit, whereas g is defined on the actual next period distribution after the uncertainty has been resolved). Recall that for each $\mathcal{J}' \subseteq \mathcal{J} \cup \{e\}$, the probability of this transition is given by $Prob\{\mathcal{J} \to \mathcal{J}'\}$ defined in 35 and the mapping between $\vec{\hat{M}}'$ and $\vec{M}'$ is given by $F_{\mathcal{J},\mathcal{J}'} : \vec{\hat{M}}' \to \vec{M}'$ defined in 34:

$$F_{\mathcal{J},\mathcal{J}'}(\vec{\hat{M}}') = \begin{cases} M'(\mathcal{I}) = \sum_{\{\mathcal{I}\in\mathcal{P}(\mathcal{J}):\mathcal{I}\cap\mathcal{J}'=\mathcal{I}'\}} \hat{M}(\mathcal{I}) & \text{, for } \mathcal{I}' \subseteq \mathcal{J} \\ M'(\mathcal{I}) = 0 & \text{, for } \mathcal{I}' \not\subseteq \mathcal{J} \end{cases}$$

Going in the reverse order, each $\mathcal{I} \in \mathcal{P}(\mathcal{J})$ is associated to $\mathcal{I}' = \mathcal{I}\cap\mathcal{J}' \in \mathcal{P}(\mathcal{J}')$; therefore $\frac{\partial g\left(F_{\mathcal{J},\mathcal{J}'}(\vec{\hat{M}}')\right)}{\partial \hat{M}'(\mathcal{I})} = \frac{\partial g(\vec{M}')}{\partial \hat{M}'(\mathcal{I})} = \frac{\partial g(\vec{M}')}{\partial M'(\mathcal{I}\cap\mathcal{J}')}\frac{\partial M'(\mathcal{I}\cap\mathcal{J}')}{\partial \hat{M}'(\mathcal{I})} = \frac{\partial g(\vec{M}')}{\partial M'(\mathcal{I}\cap\mathcal{J}')}$.
Then, we can apply this to the result of the case without entry and exit and we have:

$$\frac{\partial g\left(F_{\mathcal{J},\mathcal{J}'}(\vec{\hat{M}}')\right)}{\partial \rho_j} = \sum_{\mathcal{I}\in\mathcal{P}_{-j}} M(\mathcal{I}) \sum_{\mathcal{I}'\in\mathcal{P}_{-j},\mathcal{I}'\supseteq\mathcal{I}} \frac{\Theta(\mathcal{I}\to\mathcal{I}')}{1-\rho_j}\left[\frac{\partial g}{\partial M'((\mathcal{I}'\cup\{j\})\cap\mathcal{J}')} - \frac{\partial g}{\partial M'(\mathcal{I}'\cap\mathcal{J}')}\right]$$

Note that if $j \notin \mathcal{J}'$, then this derivative is 0 since in this case $\frac{\partial g}{\partial M'((\mathcal{I}'\cup\{j\})\cap\mathcal{J}')} = \frac{\partial g}{\partial M'(\mathcal{I}'\cap\mathcal{J}')}$.
With this, the expected value is defined as:

$$\mathbb{E}g\left(\vec{M}\right) = \sum_{\mathcal{J}'\subseteq\mathcal{J}\cup\{e\}} Prob\{\mathcal{J}\to\mathcal{J}'\} g\left(F_{\mathcal{J},\mathcal{J}'}(\vec{\hat{M}}')\right)$$

Proof of the Proposition 1: For the informative motive it is straightforward from applying 1 of Lemma 1, together with the note on the Uncertainty case:

$$\frac{\partial \mathbb{E}V_j(\mathcal{J}',\vec{M}')}{\partial \rho_j}\frac{\partial \rho_j}{\partial e_j} = \left(\sum_{\mathcal{J}'\subseteq\mathcal{J}\cup\{e\}} Prob\{\mathcal{J}\to\mathcal{J}'\} \sum_{\mathcal{I}\in\mathcal{P}_{-j}} M(\mathcal{I}) \sum_{\mathcal{I}'\in\mathcal{P}_{-j},\mathcal{I}'\supseteq\mathcal{I}} \frac{\Theta(\mathcal{I}\to\mathcal{I}')}{1-\rho_j}\left[\frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'((\mathcal{I}'\cup\{j\})\cap\mathcal{J}')} - \frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cap\mathcal{J}')}\right]\right)\frac{\partial \rho_j}{\partial e_j} > 0$$

where the positive comes from the fact that $\frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'((\mathcal{I}'\cup\{j\})\cap\mathcal{J}')} \geq \frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cap\mathcal{J}')}$, since firm j's value increases more if we add a consumer that besides $\mathcal{I}'\cap\mathcal{J}'$ she is also aware of j (there is equality if j has exited in the scenario with $\mathcal{J}'$). For the result that the informative motive decreases if we add $\{j\}$ to some consumers that weren't aware, note that the above expression is a summation over the awareness sets that don't contain j, and the change described implies a reduction of the masses in these sets.

For the anticompetitive motive, I use 2 of 1, which together with the note on the Uncertainty case, implies:

$$\sum_{k\neq j}\left(-\frac{\partial \mathbb{E}V_j(\mathcal{J}',\vec{M}')}{\partial \rho_{j'}}\right)\left(-\frac{\partial \rho_k}{\partial e_j}\right) =$$

$$\sum_{k\neq j}\sum_{\mathcal{J}'\subseteq\mathcal{J}\cup\{e\}} Prob\{\mathcal{J}\to\mathcal{J}'\}\left(\sum_{\mathcal{I}\in\mathcal{P}_{-k,-j}} M(\mathcal{I}) \sum_{\substack{\mathcal{I}'\in\mathcal{P}_{-k,-j}\\ \mathcal{I}'\supseteq\mathcal{I}}} \frac{\Theta(\mathcal{I}\to\mathcal{I}')}{1-\rho_k}\underbrace{\left[\left[\frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}')} - \frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cup\{k\})}\right] - \left[\frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cup\{j\})} - \frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cup\{j,k\})}\right]\right]}_{<0}\right.$$

$$\left. + \sum_{\mathcal{I}\in\mathcal{P}_{-k,-j}} (M(\mathcal{I}) + M(\mathcal{I}\cup\{j\})) \sum_{\substack{\mathcal{I}'\in\mathcal{P}_{-k,-j}\\ \mathcal{I}'\supseteq\mathcal{I}}} \frac{\Theta(\mathcal{I}\cup\{j\}\to\mathcal{I}'\cup\{j\})}{1-\rho_k}\underbrace{\left[\frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cup\{j\})} - \frac{\partial V_j(\mathcal{J}',\vec{M}')}{\partial M'(\mathcal{I}'\cup\{j,k\})}\right]}_{>0}\right)\left(-\frac{\partial \rho_k}{\partial e_j}\right) > 0$$

where the negative sign of the first underbrace is due to $\left(\frac{\partial V_j}{\partial M'(\mathcal{I}\cup\{j\})} - \frac{\partial V_j}{\partial M'(\mathcal{I}\cup\{j\}\cup\{j'\})}\right) > \left(\frac{\partial V_j}{\partial M'(\mathcal{I})} - \frac{\partial V_j}{\partial M'(\mathcal{I}\cup\{j'\})}\right)$ (i.e. the firm value clearly is more affected if it is a customer who learns about another good (since she will reduce the spending in j) rather than if it is a non-customer who learns about another good). And the positive sign in the second underbrace is because the value of a firm decreases if a customer learns about another good. So, we see that the anticompetitive motive increases more if $M(\mathcal{I} \cup \{j\})$ increases rather than $M(\mathcal{I})$ (direct from taking the derivatives with respect to $M(\mathcal{I} \cup \{j\})$ and $M(\mathcal{I})$ in the previous expression. Finally, that the anticompetitive motive is positive follows from observing that $\Theta(\mathcal{I} \to \mathcal{I}') = (1-\rho_j)\Theta(\mathcal{I} \cup \{j\} \to \mathcal{I}' \cup \{j\}) < \Theta(\mathcal{I} \cup \{j\} \to \mathcal{I}' \cup \{j\})$, and then the negative part of the first line is offset by the second line.

The result for the persuasive motive is straightforward from observing that 33 is a summation over the awareness sets that contain j.

6.8 Extension: Different definition of the demand shifter. Persuasive effect no harm to the taste of other varieties

Here, $\omega_{j,i,t} = 1 + \nu_s(T_t\alpha_{j,i,t})^{\nu_c}$. With this, now the derivatives with respect to $e_{j,i,t}$ write:

6.9 Social planner problem

$$\max_{\{N_{M,t},N_{j,i,t},h_{e,i,t},p_{j,i,t},\alpha_{j,i,t}\}} U = \sum_{t=0}^{\infty} \beta^t \int_0^1 [\ln C_{\ell t} + L_{\ell t}]\, d\ell$$

s.t. $C_{\ell,t}$ from (5), $C_{\ell,i,t}$ from (6), $c_{\ell,j,i,t}$ from (11), and $L_{\ell,t}$ from 4, with $T_t = Q_t$

$$y_{j,i,t} = N_{j,i,t}, \quad y_{0,i,t} = A_0 N_{0,i,t}, \quad Q_t = AN_{m,t}^{\varphi} \quad \text{(Production functions)}$$

$$1 = N_{m,t} + \int_0^1 \left(\sum_{j\in\{0\}\cup\mathcal{J}_{i,t}} N_{j,i,t} + N_{e,i,t} \right) di \quad , \quad w_t = E_t \quad \text{(Resource constraints)}$$

$$\sum_{j\in\mathcal{J}_{i,t}} \alpha_{j,i,t} = \alpha_i, (9), (10), (14), (34), (35) \quad \text{(Learning process)}$$

$$z_{e,i,t} = \phi N_{e,i,t}^{\frac{1}{2}}, (34), (35) \quad \text{(Entry and exit)}$$

Plugging $C_{\ell,t}$ and $L_{\ell,t}$ with $T_t = Q_t$ into the objective function and interchanging the integrals over ℓ and i:

$$\max_{\{N_{M,t},N_{e,i,t},N_{j,i,t},p_{j,i,t},\alpha_{j,i,t}\}} U = \int_0^1 \int_0^1 \sum_{t=0}^{\infty} \beta^t \ln C_{\ell,i,t} d\ell di + \sum_{t=0}^{\infty} \beta^t \upsilon \frac{Q_t^2}{2}$$

s.t. $C_{\ell,i,t}$ from (6), $c_{\ell,j,i,t}$ from (11)

$$y_{j,i,t} = N_{j,i,t}, \quad y_{0,i,t} = A_0 N_{0,i,t}, \quad Q_t = AN_{m,t}^{\varphi} \quad \text{(Production functions)}$$

$$1 = N_{m,t} + \int_0^1 \left(\sum_{j\in\{0\}\cup\mathcal{J}_{i,t}} N_{j,i,t} + N_{e,i,t} \right) di \quad , \quad w_t = E_t \quad \text{(Resource constraints)}$$

$$\sum_{j\in\mathcal{J}_{i,t}} \alpha_{j,i,t} = \alpha_i, (9), (10), (14), (34), (35) \quad \text{(Learning process)}$$

$$z_{e,i,t} = \phi N_{e,i,t}^{\frac{1}{2}}, (34), (35) \quad \text{(Entry and exit)}$$

The planner decides how much to produce for each individual and, accordingly, sets the prices that induce the consumers to consume these quantities. Let N_t^P be the labor used to produce all the goods in the production sector. Then, The FOC for $c_{\mathcal{I},j,i,t}$ writes:

$$[c_{\mathcal{I},j,i,t}]: \quad \beta^t \frac{1}{C_t} \frac{\partial C_t}{\partial C_{i,t}} \frac{\partial C_{i,t}}{\partial C_{\mathcal{I},i,t}} \frac{\partial C_{\mathcal{I},i,t}}{\partial c_{\mathcal{I},j,i,t}} = \beta^t \lambda \frac{\partial N_t^P}{\partial C_t} \frac{\partial C_t}{\partial C_{i,t}} \frac{\partial C_{i,t}}{\partial C_{\mathcal{I},i,t}} \frac{\partial C_{\mathcal{I},i,t}}{\partial c_{\mathcal{I},j,i,t}} \tag{36}$$

1. Dividing both sides $\frac{\partial C_t}{\partial C_{i,t}} \frac{\partial C_{i,t}}{\partial C_{\mathcal{I},i,t}} \frac{\partial C_{\mathcal{I},i,t}}{\partial c_{\mathcal{I},j,i,t}} > 0$, and defining $\hat{P}_t = \frac{w_t N_t^P}{C_t}$, we get:

$$\frac{1}{C_t} = \lambda \frac{\hat{P}_t}{w_t} \implies \lambda = \frac{w_t}{\hat{P}_t C_t} \tag{37}$$

2. $\ln C_t = \int_0^1 \ln C_{i,t} di$. Dividing both sides of 36 by $\frac{\partial C_{i,t}}{\partial C_{\mathcal{I},i,t}} \frac{\partial C_{\mathcal{I},i,t}}{\partial c_{\mathcal{I},j,i,t}} > 0$, letting $N_{i,t}$ be the the labor used in sector i, $\frac{\partial N_t^P}{\partial C_t} \frac{\partial C_t}{\partial C_{i,t}} = \frac{\partial N_t^P}{\partial C_{i,t}}$ and defining $\hat{P}_{i,t} = \frac{w_t N_{i,t}}{C_{i,t}}$, we get:

$$\frac{1}{C_{i,t}} = \lambda \frac{\hat{P}_{i,t}}{w_t} \implies \lambda = \frac{w_t}{\hat{P}_{i,t} C_{i,t}} \implies C_{i,t} = C_t \frac{\hat{P}_t}{\hat{P}_{i,t}} \tag{38}$$

where for the last expression I have used 37. Plugging $C_{i,t}$ into the definition of C_t, we get:

$$\ln \hat{P}_t = \int_0^1 \ln \hat{P}_{i,t} di \tag{39}$$

3. $\ln C_{i,t} = \int_0^1 \ln C_{\ell,i,t} d\ell$. Dividing both sides of 36 by $\frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} > 0$, letting $N_{\ell,i,t}$ be the the labor used in sector i by ℓ, $\frac{\partial N_{i,t}}{\partial C_{i,t}} \frac{\partial C_{i,t}}{\partial C_{\ell,i,t}} = \frac{\partial N_{i,t}}{\partial C_{\ell,i,t}}$ and defining $\hat{P}_{\ell,i,t} = \frac{w_t N_{\ell,i,t}}{C_{\ell,i,t}}$, we get:

$$\frac{1}{C_{\ell,i,t}} = \lambda \frac{\hat{P}_{\ell,i,t}}{w_t} \implies \lambda = \frac{w_t}{\hat{P}_{\ell,i,t} C_{\ell,i,t}} \implies C_{\ell,i,t} = C_{i,t} \frac{\hat{P}_{i,t}}{\hat{P}_{\ell,i,t}} \tag{40}$$

where for the last expression I have used 38. Plugging $C_{\ell,i,t}$ into the definition of $C_{i,t}$, we get:

$$\ln \hat{P}_{i,t} = \int_0^1 \ln \hat{P}_{\ell,i,t} di \tag{41}$$

4. $C_{\ell,i,t}$ given by 6. Letting $N_{\ell,j,i,t}$ be the the labor used in good j in sector i by ℓ, $\frac{\partial N_{\ell,i,t}}{\partial C_{\ell,i,t}} \frac{\partial C_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = \frac{\partial N_{\ell,i,t}}{\partial c_{\ell,j,i,t}} = \frac{1}{A_j}$, we get:

$$\frac{1}{C_{\ell,i,t}} \left(\frac{C_{\ell,i,t}}{c_{\ell,j,i,t}} \right)^{\frac{1}{\sigma}} \omega_{j,i,t} = \lambda \frac{1}{A_j} \implies \left(\frac{C_{\ell,i,t}}{c_{\ell,j,i,t}} \right)^{\frac{1}{\sigma}} \omega_{j,i,t} = \frac{w_t}{\hat{P}_{\ell,i,t} A_j} \implies c_{\ell,j,i,t} = C_{\ell,i,t} \hat{P}^{\sigma}_{\mathcal{I},i,t} \left(\omega_{j,i,t} \frac{A_j}{w_t} \right)^{\sigma} \tag{42}$$

where I have used λ from 40. Plugging $c_{\ell,j,i,t}$ into the definition of $C_{\ell,i,t}$, we get:

$$\hat{P}_{\mathcal{I},i,t} = \left(\left(\frac{A_0}{w_t} \right)^{\sigma-1} + \sum_{j \in \mathcal{I}} \omega^{\sigma}_{j,i,t} \left(\frac{1}{w_t} \right)^{\sigma-1} \right)^{\frac{1}{1-\sigma}} \tag{43}$$

Since we have $\hat{P}_{\ell,i,t} C_{\ell,i,t} = \hat{P}_{i,t} C_{i,t} = \hat{P}_t C_t = w_t N^P_t$ and λ from 37, then we have:

$$c_{\mathcal{I},j,i,t} = w_t N^P_t \hat{P}^{\sigma-1}_{\mathcal{I},i,t} \left(\omega_{j,i,t} \frac{A_j}{w_t} \right)^{\sigma}, \quad N^P_t = \frac{1}{\lambda} \tag{44}$$

Comparing this with the consumer choices:

$$c_{\mathcal{I},j,i,t} = E_t P^{\sigma-1}_{\mathcal{I},i,t} p^{-\sigma}_{j,i,t} \omega^{\sigma}_{j,i,t}, \quad P_{\mathcal{I},i,t} = \left(p^{1-\sigma}_{0,i,t} + \sum_{j \in \mathcal{I}} \omega^{\sigma}_{j,i,t} p^{1-\sigma}_{j,i,t} \right)^{\frac{1}{1-\sigma}}$$

It is straightforward to check that the planner can induce the consumer to consume the quantities in 44 by setting prices equal to the marginal cost times a markup (or a tax) equal to the ratio of expenditure to the production costs; i.e. $p_{j,i,t} = \frac{w_t}{A_j} \tau_t$, where $\tau_t = \frac{E_t}{w_t N^P_t}$.

The particular level of τ affects the level of consumption, but not the share of expenditure allocated to each good, since, as seen in the following expression, $s_{\mathcal{I},j,i,t}$ is independent of τ (that is, τ doesn't distort how N^P_t is allocated among the production goods):

$$s_{\mathcal{I},j,i,t} = \frac{p_{j,i,t} c_{\mathcal{I},j,i,t}}{E_{\mathcal{I},i,t}} = \frac{\tau \frac{w_t}{A_j} c_{\mathcal{I},j,i,t}}{\tau w_t N_{\mathcal{I},i,t}} = \omega^{\sigma}_{j,i,t} \left(\frac{w_t}{A_j \hat{P}_{\mathcal{I},i,t}} \right)^{1-\sigma} \implies s_{\mathcal{I},j,i} = \frac{\omega^{\sigma}_{j,i}}{A_0^{\sigma-1} + \sum_{k \in \mathcal{I}} \omega^{\sigma}_{k,i}} \tag{45}$$

Using that $c_{\mathcal{I},j,i} = A_j N_{\mathcal{I},j,i} = A_j \frac{N_{\mathcal{I},j,i}}{N_{\mathcal{I},i}} N_{\mathcal{I},i} = A_j s_{\mathcal{I},j,i} \frac{E_t}{w_t} \frac{1}{\tau}$; then, we can write $C_{\mathcal{I},i}$ as:

$$C_{\mathcal{I},i,t} = \frac{1}{\tau} \frac{E_t}{w_t} \left((A_0 s_{\mathcal{I},0,i,t})^{\frac{\sigma-1}{\sigma}} + \sum_{k \in \mathcal{I}} \omega_{k,i,t} (s_{\mathcal{I},k,i,t})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{46}$$

and combining this with $\frac{E_t}{\tau} = w_t N^P_t = C_{\mathcal{I},i,t} \hat{P}_{\mathcal{I},i,t}$, we get (where for the second equality I use 45):

$$\left((A_0 s_{\mathcal{I},0,i,t})^{\frac{\sigma-1}{\sigma}} + \sum_{k\in\mathcal{I}} \omega_{k,i,t}(s_{\mathcal{I},k,i,t})^{\frac{\sigma-1}{\sigma}} \right) = \left(\frac{w_t}{\hat{P}_{\mathcal{I},i,t}} \right)^{\frac{\sigma-1}{\sigma}} = \omega_{j,i,t} s_{\mathcal{I},j,i,t}^{-\frac{1}{\sigma}} \tag{47}$$

Next, I move to the advertising part of the planner problem. We will use the following derivatives:

$$\frac{\partial C_{\mathcal{I},i,t}}{\partial \omega_{j,i}} = \frac{\sigma}{\sigma-1} C_{\mathcal{I},i,t}^{\frac{1}{\sigma}} \left[s_{\mathcal{I},j,i,t}^{\frac{\sigma-1}{\sigma}} + \frac{\sigma-1}{\sigma} \left(\sum_{k\in\mathcal{I}} \omega_{k,i,t} s_{\mathcal{I},k,i,t}^{\frac{-1}{\sigma}} \frac{\partial s_{\mathcal{I},k,i,t}}{\partial \omega_{j,i}} + A_0^{\frac{\sigma-1}{\sigma}} s_{\mathcal{I},0,i,t}^{\frac{-1}{\sigma}} \frac{\partial s_{\mathcal{I},0,i,t}}{\partial \omega_{j,i}} \right) \right]$$

$$\frac{\partial s_{\mathcal{I},j,i,t}}{\partial \omega_{j,i}} = s_{\mathcal{I},j,i,t}(1 - s_{\mathcal{I},j,i,t}) \frac{\sigma}{\omega_{j,i,t}}, \qquad \frac{\partial s_{\mathcal{I},k,i,t}}{\partial \omega_{j,i}} = -s_{\mathcal{I},j,i,t} s_{\mathcal{I},k,i,t} \frac{\sigma}{\omega_{j,i,t}}$$

$$\frac{\partial \omega_{j,i}}{\partial \alpha_{j,i,t}} = \nu_c \nu_s T^{\nu_c} \alpha_{j,i,t}^{\nu_c - 1} = \frac{T}{\alpha_{j,i,t}} \frac{\partial \omega_{j,i}}{\partial T}$$

The term in parenthesis of the first line can be rewritten as (in the second expression, I use 47):

$$\left(\omega_{j,i,t} - \sum_{k\in\mathcal{I}} \omega_{k,i,t} s_{\mathcal{I},k,i,t}^{\frac{\sigma-1}{\sigma}} - A_0^{\frac{\sigma-1}{\sigma}} s_{\mathcal{I},0,i,t}^{\frac{\sigma-1}{\sigma}} \right) s_{\mathcal{I},j,i,t} \frac{\sigma}{\omega_{j,i,t}} = \left(\omega_{j,i,t} - \omega_{j,i,t} s_{\mathcal{I},j,i,t}^{-\frac{1}{\sigma}} \right) s_{\mathcal{I},j,i,t} \frac{\sigma}{\omega_{j,i,t}} = \left(1 - s_{\mathcal{I},j,i,t}^{-\frac{1}{\sigma}} \right) s_{\mathcal{I},j,i,t} \sigma < 0$$

so, the term in the parenthesis is negative. And we have:

$$\frac{\partial \ln C_{\mathcal{I},i,t}}{\partial \omega_{j,i}} = \frac{\sigma}{\sigma-1} C_{\mathcal{I},i,t}^{\frac{1-\sigma}{\sigma}} \left[s_{\mathcal{I},j,i,t}^{\frac{\sigma-1}{\sigma}} + (\sigma-1)\left(1 - s_{\mathcal{I},j,i,t}^{-\frac{1}{\sigma}}\right) s_{\mathcal{I},j,i,t} \right] = \frac{\sigma}{\sigma-1} \left(\frac{s_{\mathcal{I},j,i,t}}{C_{\mathcal{I},i,t}} \right)^{\frac{\sigma-1}{\sigma}} \left[1 + (\sigma-1)\left(s_{\mathcal{I},j,i,t}^{\frac{1}{\sigma}} - 1 \right) \right]$$

So:

$$\frac{\partial \ln C_{\mathcal{I},i,t}}{\partial \alpha_{j,i}} = \left(\frac{s_{\mathcal{I},j,i,t}}{C_{\mathcal{I},i,t}} \right)^{\frac{\sigma-1}{\sigma}} \left[\frac{\sigma}{\sigma-1} + \sigma \left(s_{\mathcal{I},j,i,t}^{\frac{1}{\sigma}} - 1 \right) \right] \nu_c \nu_s T^{\nu_c} \alpha_{j,i,t}^{\nu_c - 1}$$

For the dynamic problem of advertising/media, it is useful to define $U_X = \int_0^1 \sum_{t=0}^{\infty} \beta^t \ln C_{\ell,i,t} d\ell$ as the expected life-time industry-consumption utility of an industry with the current industry state being X.

The social planner has to decide on (i) how much labor to allocate to the media sector, $N_{m,t}$, and (ii) how to allocate the ad space among the differentiated firms of each industry, $\alpha_{j,i,t}$.

First, let's see the social planner choice of $\alpha_{j,i,t}$. The allocation of the ad space has to be such that the marginal social gain of increasing the ad space given to each firm is the same, since otherwise we could improve the allocation. Formally, it must be $\beta \frac{\partial \mathbb{E} U_{X'}}{\partial \rho_{j,i}} \frac{\partial \rho_{j,i}}{\partial \alpha_{j,i}} + \frac{\partial \ln C_{X,t}}{\partial \alpha_{j,X}} = \hat{h}_X$ for some $\hat{h}_X$ and all $j \in \mathcal{J}_X$, together with $\sum\limits_{j\in\mathcal{J}_X} \alpha_{j,X} = \alpha_X$.

Second, let's see the social planner choice of N_m.

$$\left[\frac{\partial L}{\partial Q} + \sum_{X\in\Omega} \mu_t(X) \sum_{j\in\mathcal{J}_X} \left[\beta \frac{\partial \mathbb{E} U_{X'}}{\partial \rho_{j,X}} \frac{\partial \rho_{j,X}}{\partial T} + \frac{\partial \ln C_{X,t}}{\partial \omega_{j,X}} \frac{\partial \omega_{j,X}}{\partial T} \right] \frac{\partial T}{\partial Q} \right] \frac{\partial Q}{\partial N_m} = \lambda$$

where $\frac{\partial L}{\partial \bar{Q}} = \upsilon Q$, $\frac{\partial T}{\partial Q} = 1$ (if $Q < 1$, otherwise it is 0). Also, using that $\frac{\partial \rho_{j,X}}{\partial T} = \frac{\alpha_{j,X}}{T}\frac{\partial \rho_{j,X}}{\partial \alpha_{j,X}}$, $\frac{\partial \omega_{j,X}}{\partial T} = \frac{\alpha_{j,X}}{T}\frac{\partial \omega_{j,X}}{\partial \alpha_{j,X}}$, and $\frac{\partial Q}{\partial N_m} = \varphi \frac{Q}{N_m}$

$$\left[\upsilon Q + \sum_{X\in\Omega} \mu_t(X) \sum_{j\in\mathcal{J}_X} \left[\beta \frac{\partial \mathbb{E}U_{X'}}{\partial \rho_{j,X}} \frac{\partial \rho_{j,X}}{\partial \alpha_{j,X}} + \frac{\partial \ln C_{X,t}}{\partial \alpha_{j,X}}\right] \frac{\alpha_{j,X}}{T}\right] \varphi \frac{Q}{N_m} = \lambda$$

and using that $\frac{\partial \mathbb{E}U_{X'}}{\partial \rho_{j,X}} \frac{\partial \rho_{j,X}}{\partial \alpha_{j,X}} + \frac{\partial \ln C_{X,t}}{\partial \alpha_{j,X}} = \hat{h}_X$ for some value $\hat{h}_X$ and all j, that $\sum_j \alpha_{j,X} = \alpha_X$, and $Q = T$, then the condition for N_m writes:

$$\upsilon Q^2 + \sum_{X\in\Omega} \mu(X) \hat{h}_X \alpha_X = \frac{\lambda}{\varphi} N_m \tag{48}$$

Finally, the labor employed in entry in each industry satisfies:

$$\lambda = \frac{\phi}{2} N_{e,X}^{-\frac{1}{2}} \beta \left(\mathbb{E}_e U_{X'} - \mathbb{E}_{-e} U_{X'}\right) \tag{49}$$

where $\mathbb{E}_e U_{X'}$ (resp. $\mathbb{E}_{-e} U_{X'}$) is the expected industry-utility conditional on successfully creating (resp. not creating) a new differentiated good (so the expectation comes from the probabilities the incumbents exit).
Using 44, 49 and 48, the labor market clearing condition writes:

$$1 = N^P + N_e + N_m \implies \lambda = 1 + \upsilon Q^2 \varphi + \sum_{X\in\Omega} \mu(X) \left(\varphi \hat{h}_X \alpha_X + \left(\frac{\phi}{2}\beta \left(\mathbb{E}_e U_{X'} - \mathbb{E}_{-e} U_{X'}\right)\right)^2 \lambda^{-1}\right) \tag{50}$$

Note that this clearly implies $\lambda > 1$. Finally, the budget constraint implies the relative wage is 1, $\hat{w} = \frac{w}{E} = 1$. Therefore the planner's markup (or tax) is $\tau = \frac{1}{N^P \hat{w}} = \frac{1}{N^P} = \lambda > 1$.

6.10 Proof of convergence to an ergodic distribution and uniqueness

Uniqueness:

Let τ be the first period that we arrive at state $\mathcal{J} = \emptyset$, and $P_{t,0}(X)$ be he probability that we are at X after t periods starting from $\mathcal{J} = \emptyset$; then the probability we are at state X starting from a given state is:

$P_t\{X\} = \sum_{k=1}^{t} P\{\tau = k\} P_{t-k,0}\{X\} + P\{\tau > t\} P_t\{X|\tau > K\}$

As $t \to \infty$, $P\{\tau > t\} \to 0$ since every period there is a positive probability that all differentiated firms die and we arrive at $\mathcal{J} = \emptyset$. Therefore, this tells us that if $P_{t,0}\{X\}$ converges (which later I prove that this is the case), then, the only stationary distribution we can have is $P_0(X) = \lim_{t\to\infty} P_{t,0}(X)$.

The set of possible states is at most countably infinite

This is a consequence of two things: (i) from a given state you can directly move to a finite number of states; (ii) with probability 1 any industry will pass through the state $\mathcal{J} = \emptyset$ at some point in time. Just as in the proof of Uniqueness, (ii) is telling us that the only stationary distribution we can have (if any, since I haven't proved this yet) is the one we would converge to starting from the state $\mathcal{J} = \emptyset$, which (i) tells us that at most will have a countably infinite number of different states.

Convergence (Existence)

Suppose there are $n \in \mathbb{N} \cup \{\infty\}$ possible states and the probability of moving from state j to state i is $a_{i,j}$, then the transition matrix is

$$Q = \begin{pmatrix} 1 - \sum_{j=2}^{n} a_{j,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & 1 - \sum_{j \neq 2} a_{j,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & 1 - \sum_{j=1}^{n} a_{n,n} \end{pmatrix}$$

Let $m_t = (m_{1,t}, \ldots, m_{n,t})$ be the vector of masses in each state, and call $M_t := m_{t+1} - m_t = (Q - \mathbb{I}_n)m_t$; so $M_{i,t} = \sum_{k \neq i} m_{k,t} a_{i,k} - m_{i,t} \sum_{k \neq i} a_{k,i}$.

Lemma 2 $\sum_{k=1}^{n} M_{k,t} = 0$

Proof. Given that $M_{i,t} = \sum_{k \neq i} m_{k,t} a_{i,k} - m_{i,t} \sum_{k \neq i} a_{k,i}$; then
$\sum_{i=1}^{n} M_{i,t} = \sum_{i=1}^{n} \left[\sum_{k \neq i} m_{k,t} a_{i,k} - m_{i,t} \sum_{k \neq i} a_{k,i} \right] = \sum_{i=1}^{n} \sum_{k \neq i} m_{k,t} a_{i,k} - \sum_{i=1}^{n} \sum_{k \neq i} m_{i,t} a_{k,i}$
$= \sum_{i=1}^{n} \sum_{k \neq i} m_{k,t} a_{i,k} - \sum_{k=1}^{n} \sum_{i \neq k} m_{k,t} a_{i,k}$.
So, we just need to see that $\{k \neq i | i, k \in \{1, \ldots, n\}\} = \{i \neq k | i, k \in \{1, \ldots, n\}\}$, which is clearly satisfied by symmetry of the $\neq$-relationship.

■

And the following lemma expresses M_{t+q} for $q \in \mathbb{N}$ in terms of M_t:

Lemma 3 *For any $q \in \mathbb{N}$, $M_{t+q} = Q^q M_t$, with $M_{i,t+q} = \left(1 - \sum_{k \neq i} a_{k,i}^{(q)}\right) M_{i,t} + \sum_{k \neq i} a_{i,k}^{(q)} M_{k,t}$, where $a_{i,k}^{(q)}$ is the probability of moving from k to i in q periods.*

Proof. By definition, $M_{t+q} = (Q - \mathbb{I}_n)m_{t+q} = (Q - \mathbb{I}_n)Q^q m_t = (Q^{q+1} - Q^q)m_t = Q^q(Q - \mathbb{I}_n)m_t = Q^q M_t$.

■

My main goal here is to study the convergence of M_t towards the null vector; so, we want to establish some result that compares M_t to M_{t+q} for some $q \in \mathbb{N}$. Since M_t is an n-dimensional object, it is important to specify under which metric. To see the importance of this, let's see a counterexample that shows that not necessarily each component of M_t has to monotonically decrease in absolute value:

Lemma 4 *It is not necessarily true that* $|m_{t+1}(k) - m_t(k)| \geq |m_{t+2}(k) - m_{t+1}(k)|$ *for all* k.

Proof. Suppose that m_t is only non-zero in position i, where $m_{i,t} = 1$. Then

$$M_t = (Q - \mathbb{I}_n)m_t = \begin{pmatrix} a_{1,i} \\ \vdots \\ -\sum_{k\neq i} a_{k,i} \\ \vdots \\ a_{n,i} \end{pmatrix}, \quad m_{t+2} - m_{t+1} = \begin{pmatrix} B_1 \\ \vdots \\ B_i \\ \vdots \\ B_n \end{pmatrix}$$

where $B_j = a_{j,1}\left(1 - \sum_{k\neq i} a_{k,i} - \sum_{k\neq j} a_{k,j}\right) + \sum_{k\notin\{i,j\}} a_{j,k}a_{k,1}$ for $j \neq i$ and $B_i = -\left(\sum_{k\neq i} a_{k,i}\right)\left(1 - \sum_{k\neq i} a_{k,i}\right) + \sum_{k\neq i} a_{i,k}a_{k,i}$.

Then, we can find a counterexample by just supposing $a_{1,i} = 0$ and that there exists $k \notin \{1, i\}$ such that $a_{1,k}a_{k,i} > 0$. Then, we have: $|m_{1,t+1} - m_{1,t}| = a_{1,i} = 0 < a_{1,k}a_{k,i} \leq \sum_{k\notin\{1,i\}} a_{1,k}a_{k,i} = |B_1| = |m_{1,t+2} - m_{1,t+1}|$

■

The norm that will prove useful is $||M_t|| := \max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} M_{k,t}|\}$. Define $\mathcal{B}^+ := \{i \in \{1,\ldots,n\} | M_{i,t} > 0\}$ and $\mathcal{B}^- := \{i \in \{1,\ldots,n\} | M_{i,t} < 0\}$

Proposition 4 *It is satisfied that* $\max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} M_{k,t+1}|\} \leq \max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} M_{k,t}|\} = \sum\limits_{k\in\mathcal{B}^+} M_{k,t}$ *Further, if* $i \in \mathcal{B}^+$, $j \in \mathcal{B}^-$ *and there exists* $\ell \in \{1,\ldots,n\}$ *such that* $a^{(q)}_{\ell,i}$, $a^{(q)}_{\ell,j} > 0$ *(*ℓ *can be equal to* i *or* j*, which means that this state has period smaller or equal than* q*), then* $\max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} M_{k,t+q}|\} < \max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} M_{k,t}|\} = \sum_{k\in\mathcal{B}^+} M_{k,t}$

Proof. First, $\max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} M_{k,t}|\} = \sum_{k\in\mathcal{B}^+} M_{k,t}$ is because $\max\limits_{\mathcal{A}\subset\{1,\ldots,n\}} \{|\sum_{k\in\mathcal{A}} [m_{t+1}(k) - m_t(k)]|\} = \max\{\sum_{k\in\mathcal{B}^+} M_{k,t}, -\sum_{k\notin\mathcal{B}^+} M_{k,t}\}$ and the fact that both terms have the same value by Lemma 2.

Now, from Lemma 3, for any given $q \in \mathbb{N}$ and $\mathcal{A} \subset \{1,\ldots,n\}$, we have:

$$\sum_{k\in\mathcal{A}} M_{k,t+q} = \sum_{k\in\mathcal{A}} \left[(1 - \sum_{j\neq k} a^{(q)}_{j,k})M_{k,t} + \sum_{j\neq k} a^{(q)}_{k,j} M_{j,t}\right]$$

And I group the terms with same $M_{i,t}$. Let's focus first on the positive terms (i.e. $M_{i,t}$ for $i \in \mathcal{B}^+$):

- If $i \in \mathcal{B}^+ \cap \mathcal{A}$, then: (i) for $k = i$ we have the term $(1 - \sum\limits_{j\neq i} a^{(q)}_{j,i})M_{i,t}$; (ii) for each $k \in \mathcal{A} \setminus \{i\}$, we have the term $a^{(q)}_{k,i} M_{i,t}$.
- If $i \in \mathcal{B}^+ \setminus \mathcal{A}$: for each $k \in \mathcal{A}$, we have the term $a^{(q)}_{k,i} M_{i,t}$.

Then, the positive terms can be written as:

$$\sum_{i\in\mathcal{B}^+\cap\mathcal{A}} M_{i,t}\left(1 - \sum_{j\neq i} a^{(q)}_{j,i} + \sum_{k\in\mathcal{A}\setminus\{i\}} a^{(q)}_{k,i}\right) + \sum_{i\in\mathcal{B}^+\setminus\mathcal{A}} M_{i,t}\left(\sum_{k\in\mathcal{A}} a^{(q)}_{k,i}\right) = \sum_{i\in\mathcal{B}^+\cap\mathcal{A}} M_{i,t}\left(1 - \sum_{k\notin\mathcal{A}} a^{(q)}_{k,i}\right) + \sum_{i\in\mathcal{B}^+\setminus\mathcal{A}} M_{i,t}\left(\sum_{k\in\mathcal{A}} a^{(q)}_{k,i}\right)$$

And using that for any i $\sum_{k\neq i} a^{(q)}_{k,i} \in [0,1]$, we have

$$\sum_{i\in\mathcal{B}^+\cap\mathcal{A}} M_{i,t}\left(1 - \sum_{k\notin\mathcal{A}} a^{(q)}_{k,i}\right) + \sum_{i\in\mathcal{B}^+\setminus\mathcal{A}} M_{i,t}\left(\sum_{k\in\mathcal{A}} a^{(q)}_{k,i}\right) \leq \sum_{i\in\mathcal{B}^+} M_{i,t} \tag{51}$$

Analogously, for the negative terms (i.e $M_{i,t}$ for $i \in \mathcal{B}^-$):

- If $i \in \mathcal{A} \cap \mathcal{B}^-$, then: (i) for $k = i$ we have the term $(1 - \sum\limits_{j \neq i} a_{j,i}^{(q)}) M_{i,t}$; (ii) for each $k \in \mathcal{A} \setminus \{i\}$, we have the term $a_{k,i}^{(q)} M_{i,t}$
- If $i \in \mathcal{B}^- \setminus \mathcal{A}$: for each $k \in \mathcal{A}$, we have the term $a_{k,i}^{(q)} M_{i,t}$

Then, the negative terms can be written as:

$$\sum_{i \in \mathcal{B}^- \cap \mathcal{A}} M_{i,t} \left(1 - \sum_{j \neq i} a_{j,i}^{(q)} + \sum_{k \in \mathcal{A} \setminus \{i\}} a_{k,i}^{(q)} \right) + \sum_{i \in \mathcal{B}^- \setminus \mathcal{A}} M_{i,t} \left(\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} \right) = \sum_{i \in \mathcal{B}^- \cap \mathcal{A}} M_{i,t} \left(1 - \sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} \right) + \sum_{i \in \mathcal{B}^- \setminus \mathcal{A}} M_{i,t} \left(\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} \right)$$

So, again using that for any i $\sum_{k \neq i} a_{k,i}^{(q)} \in [0, 1]$, we have

$$\sum_{i \in \mathcal{B}^- \cap \mathcal{A}} (-M_{i,t}) \left(1 - \sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} \right) + \sum_{i \in \mathcal{B}^- \setminus \mathcal{A}} (-M_{i,t}) \left(\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} \right) \leq - \sum_{i \in \mathcal{B}^-} M_{i,t} \tag{52}$$

The first part of the proposition follows directly from the fact that the previous two inequalities for $q = 1$ imply that for any $\mathcal{A} \subset \{1, \ldots, n\}$, we have $|\sum_{k \in \mathcal{A}} M_{k,t+1}| \leq \sum_{k \in \mathcal{B}^+} M_{k,t}$, and so the inequality is also true for the maximum.

For the second part of the proposition, suppose the condition holds and I will show by contradiction that we cannot find any $\mathcal{A}$ such that $|\sum_{k \in \mathcal{A}} M_{k,t+q}| \leq \sum_{k \in \mathcal{B}^+} M_{k,t}$ holds with equality, and so the inequality has to be strict.

In order for the equality to hold, it must be one of the following two cases:

- Case A: The positive terms are equal to the upper bound, and the negative terms are zero. For this to be the case, we need: (i) for $i \in \mathcal{B}^+ \cap \mathcal{A}$, $\sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} = 0$; (ii) for $i \in \mathcal{B}^+ \setminus \mathcal{A}$, $\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} = 1$; (iii) for $i \in \mathcal{B}^- \cap \mathcal{A}$, $\sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} = 1$; and (iv) for $\mathcal{B}^- \setminus \mathcal{A}$, it must be $\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} = 0$

  If $i \in \mathcal{A}$, then condition (i) implies that also $\ell \in \mathcal{A}$, since otherwise we would have the contradiction $0 = \sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} \geq a_{\ell,i}^{(q)} > 0$. If $i \notin \mathcal{A}$, then condition (ii) again implies that $\ell \in \mathcal{A}$, since otherwise we would have the contradiction $1 = \sum_{k \in \mathcal{A}} a_{k,i}^{(q)} \leq 1 - a_{\ell,i}^{(q)} < 0$. Therefore, ℓ must be in $\mathcal{A}$ in order for the positive terms to reach the upper bound.

  Next, if $j \in \mathcal{A}$, then condition (iii) implies that $\ell \notin \mathcal{A}$, since otherwise we would have the contradiction $1 = \sum_{k \notin \mathcal{A}} a_{k,j}^{(q)} \leq 1 - a_{\ell,j}^{(q)} < 1$. So, the only possibility is that $j \notin \mathcal{A}$, but then condition (iv) contradicts that $\ell \in \mathcal{A}$, since then we would have the contradiction $0 = \sum_{k \in \mathcal{A}} a_{k,j}^{(q)} \geq a_{\ell,j}^{(q)} > 0$. Therefore, Case A is not possible.

- Case B: The positive terms are equal to zero, and the negative terms are equal to the lower bound. For this to be the case, we need: (i) for $i \in \mathcal{B}^+ \cap \mathcal{A}$, $\sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} = 1$; (ii) for $i \in \mathcal{B}^+ \setminus \mathcal{A}$, $\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} = 0$; (iii) for $i \in \mathcal{B}^- \cap \mathcal{A}$, $\sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} = 0$; and (iv) for $\mathcal{B}^- \setminus \mathcal{A}$, it must be $\sum_{k \in \mathcal{A}} a_{k,i}^{(q)} = 1$

  Analogously as in the previous case, we get to the conclusion that this case is not possible.

  If $i \in \mathcal{A}$, then (i) implies $\ell \notin \mathcal{A}$, since otherwise $1 = \sum_{k \notin \mathcal{A}} a_{k,i}^{(q)} \leq 1 - a_{\ell,i}^{(q)} < 1$. If $i \notin \mathcal{A}$, then (ii) also implies that $\ell \notin \mathcal{A}$, since otherwise $0 = \sum_{k \in \mathcal{A}} a_{k,i}^{(q)} \geq a_{\ell,i}^{(q)} > 0$. So, it must be $\ell \notin \mathcal{A}$.

  If $j \in \mathcal{A}$, (iii) implies the contradiction $0 = \sum_{k \notin \mathcal{A}} a_{k,j}^{(q)} \geq a_{\ell,j} > 0$. But if $j \notin \mathcal{A}$, (iv) also implies the contradiction $1 = \sum_{k \in \mathcal{A}} a_{k,j}^{(q)} \leq 1 - a_{\ell,j} < 1$. So, we conclude that this case is not possible.

■

This proposition tells us that a sufficient condition to guarantee convergence to an ergodic distribution is that whenever we are not in a stationary distribution, we can find states that have changed in opposite directions in the previous iteration (period) such that there exists

some state which can be reached from each of the two states with positive probability in the same number of periods (in other words, if two points start from state i and j respectively, there is positive probability they will meet at some future period).
The next definitions and proposition show a sufficient condition for this condition to hold:

Definition 1 *A Markov chain is* ***irreducible*** *if for any pair of states i, j, there exists $q \in \mathbb{N}$ such that $a_{j,i}^{(q)} > 0$. (that is, it is possible to get to any state from any other state)*

Definition 2 *Let the* ***longitude of the shortest path between two states*** *i, j be $d_{i,j} = \min\{q \in \mathbb{N} | a_{i,j}^{(q)} > 0\}$ (that is, the smallest number of periods required to go from one state to the other).*

Proposition 5 *In an irreducible Markov chain that contains at least one state i with $d_{i,i} = 1$, as long as we are not in the stationary distribution, it is always possible to find states $j \in \mathcal{B}^+$ and $k \in \mathcal{B}^-$, and a state ℓ such that $a_{\ell,j}^{(q)}, a_{\ell,k}^{(q)} > 0$ for some $q \in \mathbb{N}$ (and so, in such Markov chain we can guarantee convergence to an ergodic distribution).*

Proof. If we are not in a stationary distribution then there are j with $M_{j,t} \neq 0$, and by Lemma 2 there must be $j \in \mathcal{B}^+$ and $k \in \mathcal{B}^-$. .Let i be the state such that $d_{i,i} = 1$. Then, it is sufficient to see that we can find $q \in \mathbb{N}$ such that $a_{i,j}^{(q)}, a_{i,k}^{(q)} > 0$, which is straightforward. We can check that $q := \max(d_{i,j}, d_{i,k})$ satisfies this (intuitively, the first to arrive from one of the two states then stays with positive probability in i, and at some point the one that started from the other state will also arrive to i). Without loss of generality, assume $\max(d_{i,j}, d_{i,k}) = d_{i,k}$. $a_{i,k}^{d_{i,k}} > 0$ by definition of $d_{i,k}$. But also $a_{i,j}^{d_{i,k}} \geq a_{i,j}^{d_{i,j}} a_{i,i}^{(d_{i,k}-d_{i,j})} \geq a_{i,j}^{d_{i,j}} \left(a_{i,i}^{(1)}\right)^{d_{i,k}-d_{i,j}} > 0$
■

So, in the Uniqueness section I proved that the only possible stationary distribution is the one we would obtain if the initial sate is $\mathcal{J} = \emptyset$ (if this converges). Now, the previous proposition tells us that $P_{t,0}(X)$ converges, since the Markov chain obtained is irreducible (if a state is possible, it means that there was positive probability of arriving to it starting from $\mathcal{J} = \emptyset$; and, from any state, there is probability 1 of eventually going back to $\mathcal{J} = \emptyset$) and the state $\mathcal{J} = \emptyset$ satisfies that the longitude of its shortest path connecting it to itself is 1 (with positive probability there will be no entrant and we stay at $\mathcal{J} = \emptyset$).

6.11 Summary of the method to solve the model

1. First, for each possible number of firms J:

    - Define the different awareness sets $\mathcal{P}_J$. There are 2^J awareness sets (think on how many different ways we can assign $\{0, 1\}$ to J variables).

- Define the N_J grid nodes we will use, $\vec{M}_n$, $n = 1, \ldots, N_J$. Each node is a vector of the mass of consumers in each awareness set. As we have seen, the solutions of the model are functions of the form $f(\mathcal{J}, \vec{M})$ on a continuous m-dimensional space, with $m = 2^J - 1$ ($\vec{M}$ is $(m+1)$-dimensional, but since the masses have to add up to 1, one is redundant). To deal with this exponentially increasing state space and alleviate the curse of dimensionality, I introduce a piecewise multivariate Newton interpolation method described in detail in section 6.12. Using this method, increasing the number of grid points leads to a better approximation, as in standard univariate methods using a grid and linear interpolation, with the advantage that the higher degree of the interpolating polynomial allows to reduce the number of necessary grid points for a given fit.[20]

  Also, note that $\mathcal{J}$ has information of the identity of the firm. Therefore, some nodes are just a reordering of firms, so I use this to avoid solving again nodes that are just a reordering of a node that has already been solved.

2. Define initial guesses for the aggregate states w and T, as well as initialize the policy functions for advertising expenditure and entry; that is, assign a value for the grid nodes $\{\{\{e_{j,n}\}_{j=1}^{J}\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$ and entry $\{\{N_{e,n}\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$.

3. Given the aggregate states:

   (a) Solve the firm problem:

      i. Given the guess of the policy functions for advertising expenditures and entry:

      - Solve the static price-setting problem. Note that this has to be updated in every iteration of the firm problem because the advertising choices affect the demand shifters $\omega_{j,i,t}$. This gives us profits and markups at each node: $\{\{\mathcal{M}_j(J, \vec{M}_n), \pi_j(J, \vec{M}_n)\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$.
      - Solve for the value function implied by the policy functions of advertising and entry and the profit function. Note that this implies solving a linear system on $\{\{V(J, \vec{M}_n)\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$.

      ii. Given the functions for markups and firm value found in the previous points, $\{\{\mathcal{M}_j(J, \vec{M}_n), V_j(J, \vec{M}_n)\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$, compute the best responses $\{\{\{e'_{j,n}, N'_{e,n}\}_{j=1}^{J}\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$ (i.e. the optimal choice keeping the competitors' choices fixed). If the difference between these best responses and the previous guess is small enough, we

[20] Using piecewise interpolation is important because increasing the degree of an interpolating polynomial doesn't necessary lead to a better approximation (Runge's phenomenon).

are done (in this case we have found a Nash equilibrium); otherwise, update the guesses and go back to (i).

(b) Solve for the unique stationary distribution given the solution of the firm problem (in particular, we need the policy functions for the advertising space $\{\{\{\alpha_{j,n}\}_{j=1}^{J}\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$ and entry $\{\{N_{e,n}\}_{n=1}^{N_J}\}_{J=1}^{\bar{J}}$ and entry in an industry with $J = 0$: $N_{e,0}$). For the details of the method, see section 6.11.1

4. Given the firm policy functions and the stationary distribution, compute the implied aggregates w and T, using 24, 23, together with $T = Q$. If the difference between the guesses and the implied values of w and T are close enough, we are done; otherwise, update the new guesses for w and T and go back to 3.

6.11.1 Method used to find the stationary distribution

The method has two parts.

1. In section 6.10, I show that the set of industry states observed in the stationary distribution is at most countably infinite; and so the stationary distribution is a discrete probability function defined on a potentially infinite set of points, and so, computationally, the set of states needs to be bounded some way. In the following I describe the approach used in the baseline to bound the set of states. As a robustness, I compare the stationary distribution obtained from this approach to the one obtained by bounding the space by a grid (that is, restricting $\vec{M}$ to take only values from a grid). The baseline approach tends to be much faster.

   (a) Given that in section 6.10 I show that the unique stationary distribution is the one we would obtain if the initial sate is $\mathcal{J} = \emptyset$, then:

   - I initialize the *List* of states with this state. For each state in the *List*, I store (1) the number of firms, (2) the vector of masses corresponding to this state, (3) the vector of ages, and (4) the probability *Prob*, which I now describe. *Prob* is the probability of going from state $\mathcal{J} = \emptyset$ to the particular state X in the shortest path from $\mathcal{J} = \emptyset$ to X. That is, for this initial state $\mathcal{J} = \emptyset$, we have $Prob = 1$.
   - To facilitate the process of looking up whether we have already encountered a state before (i.e. whether a state is already in *List*, I order the states in a library *LibraryStates* lexicographically based on (i) the number of active firms, (ii) the vector of ages, and (iii) the vector of masses. Initially, $LibraryStates = 1$.

- I also initialize $iter = 0$ and the list of states I will take as starting point in the following iteration, $NewStates_{iter}$. Initially, $NewStates_1 = 1$.
- Finally, we need the transition matrix with the probabilities of going from each state to the others. However, since this matrix is very sparse (the states are just directly connected to few others) and storing the whole matrix with all the zeros would be highly costly for memory storage (and solving the system would also be very slow), I only store the non-zero elements of the transition matrix in a *Library*, where each book contains three pieces of information (the books are ordered lexicographically based on the same order of these three pieces of information): the row in the transition matrix (i.e., state of origin), the column in the transition matrix (i.e., state of destination), and the value in this position of the matrix resulting from subtracting the transition matrix from the identity matrix. I initialize it as $Library = [1, 1, 1]$ (the third 1 is the 1 from the identity matrix).

(b) Then, as long as $NewStates_{iter+1}$ is not empty, increase iter by 1 and do the following for each state $s \in NewStates_{iter+1}$:

i. Calculate the next period vector of masses if there weren't entry/exit, and the probability of an entrant. Then, for each of the possible cases of entry/exit, letting q be the probability of the particular event of entry/exit, I do the following:

ii. Look up whether this state is already in *List*, using the order in *LibraryStates*. Here is where I bound the problem.

- If the probability *Prob* is above a threshold, then I check for an exact match (that is, they match in the three elements: J, the vector of ages, and the vector of masses (note that, although the vector of ages is not a state in the baseline firm problem, it is useful to distinguish it for the quantitative exercises).
- If the probability *Prob* is below the threshold (that is, it is a rare state), then I just check for J and the vector of ages. If there is no state in *List* matching J and the vector of ages, then we will treat this state as a new state; otherwise, I will treat it as if it were identical to the first state in *List* with the same J and ages. The intuition is that, although the vector of ages is not a sufficient statistic (because history matters), it serves as a good first approximation. The other boundary I set is on the firm age; in particular, I don't distinguish ages above a threshold (which

I set to 20 years old). The intuition is that for firms older than 20 years old very few consumers remain unaware of the firm, so the error from not distinguishing older firms is negligible.

iii. If the outcome from the previous point is that it is not a new state, then we index it by s' equal to the index of the state we have matched it to and go to (iv); else, if it is a new state, then we index it by s' equal to the current size of *List* plus one and do the following:

- Add the one of the identity matrix to *Library*; that is: add $[s', s', 1]$.
- Add the four pieces of information relative to this state in *List*. *Prob* will be equal to the *Prob* of state s time q.
- Add s' to $NewStates_{iter+1}$.

iv. Add $[s, s', -q]$ to the *Library*. If there is already an element at position $[s, s']$, then just add $-q$.

2. In the second part, we need to solve for the stationary distribution. The matrix found in the previous step is singular (note that the sum of all the elements in row s is $1 - \sum_{s'} p_{s,s'} = 1 - 1 = 0$, where $p_{s,s'}$ is the probability of moving from state s to s'). So, we need to add a new condition to have a compatible and determinate system: it is the condition that the solution must add up to 1; so, I add to *Library* $[s, 0, 1]$, for all the states s. Then, we also need the vector of independent coefficients, which again is very sparse (there is only one non-zero value), so again I store it in a library called $LibraryB = [0, 1]$.

6.12 Multivariate Newton Interpolation

First, as a recap of the univariate Newton interpolation, given $n+1$ different points (nodes) defined as a pair $(x_i, f(x_i))$ with $x_i \neq x_j$ for any $j \neq i$, then the unique n-degree polynomial that passes through these $n+1$ points expressed in the Newton basis polynomials (which are defined as $w_j(x) = \prod_{k=0}^{j-1}(x - x_k)$, $j = 1, 2, \ldots, n$ and $w_0(x) = 1$) is $P_n(x) = \sum_{j=0}^{n} a_j w_j(x)$, where the coefficients a_i are the solutions of the system (note that $w_j(x_i) = 0$ when $i < j$):

$$\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & (x_1 - x_0) & 0 & \ldots & 0 \\ 1 & (x_2 - x_0) & \prod_{k=0}^{1}(x_2 - x_k) & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_0) & \prod_{k=0}^{1}(x_n - x_k) & \ldots & \prod_{k=0}^{n-1}(x_n - x_k) \end{pmatrix}_{(n+1)\times(n+1)} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}$$

This system can be solved reducing the system iteratively. To this purpose, note that a_0 is already solved: $a_0 = f(x_0)$ so we can forget about row 1 and the rest writes:

$$\begin{pmatrix} f(x_0) \\ f(x_0) \\ \vdots \\ f(x_0) \end{pmatrix} + \begin{pmatrix} (x_1 - x_0) & 0 & \dots & 0 \\ (x_2 - x_0) & \prod_{k=0}^{1}(x_2 - x_k) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (x_n - x_0) & \prod_{k=0}^{1}(x_n - x_k) & \dots & \prod_{k=0}^{n}(x_n - x_k) \end{pmatrix}_{n\times n} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}$$

Next, note that each element of row i in the $n \times n$ matrix (the Vandermonde matrix) contains $(x_i - x_0)$, so we can divide both sides of row i by $(x_i - x_0)$ and calling $f[x_i, x_0] = \frac{f(x_i)-f(x_0)}{x_i - x_0}$ (divided difference) we obtain an analogous system as the initial one but with one dimension less:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & (x_2 - x_1) & 0 & \dots & 0 \\ 1 & (x_3 - x_1) & \prod_{k=1}^{2}(x_3 - x_k) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_1) & \prod_{k=1}^{2}(x_n - x_k) & \dots & \prod_{k=1}^{n-1}(x_n - x_k) \end{pmatrix}_{n\times n} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f[x_1, x_0] \\ f[x_2, x_0] \\ f[x_3, x_0] \\ \vdots \\ f[x_n, x_0] \end{pmatrix}$$

Now, $a_1 = f[x_1, x_0]$ is already solved; so, we repeat the procedure: we pass subtracting the $1 \cdot a_i$ of each row to the right hand side and we divide by the common factor of the left side $(x_i - x_1)$, we call $f[x_i, x_1, x_0] = \frac{f[x_i,x_0]-f[x_1,x_0]}{x_i - x_1}$, and we obtain:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & (x_3 - x_2) & 0 & \dots & 0 \\ 1 & (x_4 - x_2) & \prod_{k=2}^{3}(x_4 - x_k) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_2) & \prod_{k=2}^{3}(x_n - x_k) & \dots & \prod_{k=2}^{n-1}(x_n - x_k) \end{pmatrix}_{(n-1)\times(n-1)} \begin{pmatrix} a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f[x_2, x_1, x_0] \\ f[x_3, x_1, x_0] \\ f[x_4, x_1, x_0] \\ \vdots \\ f[x_n, x_1, x_0] \end{pmatrix}$$

Iterating, in the r-th iteration we will get:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & (x_r - x_{r-1}) & 0 & \dots & 0 \\ 1 & (x_{r+1} - x_{r-1}) & \prod_{k=r-1}^{r}(x_{r+1} - x_k) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_{r-1}) & \prod_{k=r-1}^{r}(x_n - x_k) & \dots & \prod_{k=r-1}^{n-1}(x_n - x_k) \end{pmatrix}_{(n-r+2)\times(n-r+2)} \begin{pmatrix} a_{r-1} \\ a_r \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f[x_{r-1}, x_{r-2} \dots, x_0] \\ f[x_r, x_{r-2} \dots, x_0] \\ f[x_{r+1}, x_{r-2} \dots, x_0] \\ \vdots \\ f[x_n, x_{r-2} \dots, x_0] \end{pmatrix}$$

Summarizing, the coefficients of the newton interpolation polynomial are given by the divided differences $a_i = f[x_i, x_{i-1}, \dots, x_0] = \frac{f[x_i,x_{i-2},\dots,x_0]-f[x_{i-1},x_{i-2},\dots,x_0]}{x_i - x_{i-1}}$.

We can extend this to the multivariate case as follows. Suppose we want to interpolate a function $f : \mathbb{R}^m \to \mathbb{R}$ by a polynomial of m variables and degree n.

Definition 3 ***(Generating points):*** *For each dimension $i = 1, \dots, m$, we define $n+1$ points $x_{i,k}$, $k = 0, \dots, n$. $\{\{x_{i,k}\}_{k=0}^{n}\}_{i=1}^{m}$ are called the generating points.*

Definition 4 (Multiindices): *Let $\vec{\alpha} = (\alpha_1, \dots, \alpha_m) \in \Lambda_{m,n} := \{\vec{\alpha} \in \{0, \dots, n\}^m | \sum\limits_{i=1}^{m} \alpha_i \leq n\}$, and $\vec{x}_{\vec{\alpha}} = (x_{1,\alpha_1}, \dots, x_{m,\alpha_m})$.*

The cardinal of $\Lambda_{m,n}$ (i.e. the number of different multiindices) is given by $N(m,n) = \binom{n+m}{n}$ (to see this, you can think of $1^{\alpha_0} x_1^{\alpha_1} \cdots x_m^{\alpha_m}$ with $\sum_{i=0}^{m} \alpha_i = n$, which we can transcribe as $\underbrace{1 \dots 1}_{\alpha_0} \# \underbrace{x_1 \dots x_1}_{\alpha_1} \# \dots \# \underbrace{x_m \dots x_m}_{\alpha_m}$; so the problem of finding the number of different multiindices is equivalent to finding the number of different ways we can choose m boxes from $n+m$ boxes (i.e. the position of the m hashtags), which is $\binom{n+m}{m}$).

Definition 5 (Newton polynomial): $w_{\vec{\alpha}}(\vec{x}) = \prod\limits_{i=1}^{m} \prod\limits_{k=0}^{\alpha_i - 1} (x_i - x_{i,k})$.

Definition 6 *The m-dimensional* ***Newton interpolating polynomial*** *of degree n of the function f is $p_{m,n}(\vec{x}) = \sum\limits_{\vec{\alpha} \in \Lambda_{m,n}} a_{\vec{\alpha}} w_{\vec{\alpha}}(\vec{x})$, satisfying $f(\vec{x}_{\vec{\alpha}}) = p_{m,n}(\vec{x}_{\vec{\alpha}})$, for all $\vec{\alpha} \in \Lambda_{m,n}$.*

Lemma 5 *Note that given $\vec{\beta}, \vec{\alpha} \in \Lambda_{m,n}$, if $\beta_i - 1 \geq \alpha_i$, then $w_{\vec{\beta}}(\vec{x}_{\vec{\alpha}})$ contains the term $(x_{i,\alpha_i} - x_{i\alpha_i}) = 0$.*

Corollary 1 *Then, $f(\vec{x}_{\vec{\alpha}}) = p_{m,n}(\vec{x}_{\vec{\alpha}}) = \sum\limits_{k_m=-1}^{\alpha_m - 1} \cdots \sum\limits_{k_1=-1}^{\alpha_1 - 1} \prod\limits_{s_m=0}^{k_m} (x_{m,\alpha_m} - x_{m,s_m}) \cdots \prod\limits_{s_m=0}^{k_m} (x_{1,\alpha_1} - x_{1,s_1}) a_{(k_1+1, \dots, k_m+1)}$*

To allow generality, I define:

Definition 7 *Given $\vec{\alpha} = (\alpha_1, \dots, \alpha_m)$, define:*
(i) $\vec{\alpha}^{(i,k)} = (\alpha_1, \dots, \alpha_{i-1}, k, \alpha_{i+1}, \dots, \alpha_m)$ (i.e. $\vec{\alpha}^{(i,k)}$ equals $\vec{\alpha}$ except k in position i)
(ii) $\vec{\alpha}^{(k)} = (\alpha_{k+1}, \dots, \alpha_m)$.
(iii) $\vec{\beta}^{(i,k)} = (\vec{\alpha}^{(i,k-1)}, \dots, \vec{\alpha}^{(i,0)}, \dots, \vec{\alpha}^{(m,\alpha_m - 1)}, \dots, \vec{\alpha}^{(m,0)})$, and let $\vec{\beta}^{(i,0)} = \vec{\beta}^{(i+1,\alpha_{i+1}-1)}$ and $\vec{\beta}^{(m,0)} = \emptyset$.

Definition 8 (Divided differences):
$f[\vec{\alpha}, \vec{\beta}^{(i,b)}] = \sum\limits_{k_i=b-1}^{\alpha_i - 1} \cdots \sum\limits_{k_1=-1}^{\alpha_1 - 1} \prod\limits_{s_i=b}^{k_i} (x_{m,\alpha_m} - x_{m,s_m}) \cdots \prod\limits_{s_1=0}^{k_1} (x_{1,\alpha_1} - x_{1,s_1}) a_{(k_1+1, \dots, k_i+1, \vec{\alpha}^{(i)})}$ if $\alpha_i > b$; and $f[\vec{\alpha}, \vec{\beta}^{(i,b)}] = f[\vec{\alpha}, \vec{\beta}^{(i-1,0)}]$ otherwise.

Note that by Corollary 1, and since $\vec{\alpha}^{(m)} = \emptyset$, then $f[\vec{\alpha}, \vec{\beta}^{(m,0)}] = f(\vec{x}_{\vec{\alpha}})$. The algorithm to find the coefficients $a_{\vec{\alpha}}$ is defined as follows:

1. *Start setting $i = m$ and $b = 0$.*

2. *If there is some $\vec{\alpha} \in \Lambda_{m,n}$ such that $\alpha_i > b$, then:*

   (a) *For all the $\vec{\alpha} \in \Lambda_{m,n}$ such that $\alpha_i > b$: Noting that $f[\vec{\alpha}^{(i,b)}, \vec{\beta}^{(i,b)}]$ contains all the terms of $f[\vec{\alpha}, \vec{\beta}^{(i,b)}]$ with $k_i = b - 1$, and so the remaining terms will all contain $(x_{i,\alpha_i} - x_{i,b})$; then:*

$$f[\vec{\alpha}, \vec{\beta}^{(i,b+1)}] = \frac{f[\vec{\alpha}, \vec{\beta}^{(i,b)}] - f[\vec{\alpha}^{(i,b)}, \vec{\beta}^{(i,b)}]}{x_{i,\alpha_i} - x_{i,b}}$$
$$= \sum_{k_i=b}^{\alpha_i-1} \cdots \sum_{k_1=-1}^{\alpha_1-1} \prod_{s_i=b+1}^{k_i} (x_{m,\alpha_m} - x_{m,s_m}) \cdots \prod_{s_1=0}^{k_1} (x_{1,\alpha_1} - x_{1,s_1}) a_{(k_1+1,\ldots,k_i+1,\vec{\alpha}^{(i)})}$$

   (b) *For all the $\vec{\alpha} \in \Lambda_{m,n}$ such that $\alpha_i \leq b$, then $f[\vec{\alpha}, \vec{\beta}^{(i,b+1)}] = f[\vec{\alpha}, \vec{\beta}^{(i,b)}]$ (satisfies the definition since $\alpha_i \leq b < b+1$, so $f[\vec{\alpha}, \vec{\beta}^{(i,b+1)}] = f[\vec{\alpha}, \vec{\beta}^{(i,b)}] = f[\vec{\alpha}, \vec{\beta}^{(i-1,0)}]$*

   Set $b = b + 1$, and go back to step 2.

3. *If $\alpha_i \leq b$ for all $\vec{\alpha} \in \Lambda_{m,n}$ (which is satisfied if and only if $b \leq n$), then make $f[\vec{\alpha}, \vec{\beta}^{(i-1,0)}] = f[\vec{\alpha}, \vec{\beta}^{(i,b)}]$, and set $i = i - 1$ and $b = 0$. If $i = 0$, we are done; otherwise, go back to step 2.*

All is left to do is to show that the $f[\vec{\alpha}, \vec{\beta}^{(0,0)}] = a_{\vec{\alpha}}$ for all $\vec{\alpha} \in \Lambda_{m,n}$. Given that the divided difference of $\vec{\alpha}$ just changes when we apply (2a) to it, then it is sufficient to see that in the last time that we select $\vec{\alpha}$ for (2a) it is $f[\vec{\alpha}, \vec{\beta}^{(i,b+1)}] = a_{\vec{\alpha}}$; since then it will be $f[\vec{\alpha}, \vec{\beta}^{(0,0)}] = f[\vec{\alpha}, \vec{\beta}^{(i,b+1)}] = a_{\vec{\alpha}}$.

Proof. *If we have used $a_{\vec{\alpha}}$ in (2a), it means that $\alpha_i > b$, which implies that exactly one of the following is true:*

1. *$\alpha_i > b + 1$, in which case $a_{\vec{\alpha}}$ would also be selected in the next iteration, contradicting it was the last time it was selected;*

2. *$\alpha_i = b + 1$, in which case $a_{\vec{\alpha}}$ it is the last iteration for variable i that $a_{\vec{\alpha}}$ is selected. In this case there are two possibilities:*

   - *$\alpha_k > 0$ for some $k < i$, in which case in iteration $(k, 0)$ $\vec{\alpha}$ would be selected, contradicting the hypothesis.*
   - *$\alpha_k = 0$ for all $k < i$, in which case we have:*

$$
\begin{aligned}
f[\vec{\alpha}, \vec{\beta}^{(i,b+1)}] &= \sum_{k_i=\alpha_i-1}^{\alpha_i-1} \cdots \sum_{k_1=-1}^{\alpha_1-1} \prod_{s_i=\alpha_i}^{k_i} (x_{m,\alpha_m} - x_{m,s_m}) \cdots \prod_{s_1=0}^{k_1} (x_{1,\alpha_1} - x_{1,s_1}) a_{(k_1+1,\ldots,k_i+1,\vec{\alpha}^{(i)})} \\
&= \prod_{s_i=\alpha_i}^{\alpha_i-1} (x_{m,\alpha_m} - x_{m,s_m}) \cdots \prod_{s_1=0}^{-1} (x_{1,\alpha_1} - x_{1,s_1}) a_{(0,\ldots,0,\alpha_i,\vec{\alpha}^{(i)})} \\
&= a_{(0,\ldots,0,\alpha_i,\vec{\alpha}^{(i)})} = a_{\vec{\alpha}}
\end{aligned}
$$

■