Attention and Blur Effects of Advertising with Spill-overs: Nested Consideration Logit Approach

Woohun Son*

April 4, 2025

Abstract

This paper develops a practical structural demand model—the Nested Consideration Logit (NCL) model—that disentangles two key effects of advertising: the *attention effect*, which increases the likelihood of consumer consideration, and the *blur effect*, which reduces perceived differences in product quality due to industry-wide advertising. The NCL allows for both product- and nest-level consideration probabilities to adjust desirability of product and nest. Blur is a function of the salience and volume of advertising in the model. The model incorporates *spill-over* effects of advertising within industries and frames ad competition as a public goods or public bads game, depending on market conditions. Comparative statics highlight how blur benefits lower-quality firms and reshapes price elasticities by weakening competitive pressure. The paper also establishes identification strategies for the model's parameters, especially when panel data exhibit blur effects invariance for some time periods. The model provides a practical tool for analyzing advertising effects in the presence of limited attention and information congestion given market level data and offers insights into how market dominance and advertising jointly shape consumer choice.

Keywords: advertising, spill-over, nested logit model, attention, information

JEL: M37, D83, L13, D12, C35

*Address: Department of Economics, The Ohio State University, 1945 N High Street, Columbus, OH 43201, USA, e-mail: son.268@osu.edu

Contents

1 Introduction **3**

2 Literature Review **5**

3 Nested Consideration Logit Model **12**

3.1 Comparative Statics: Attention effect of advertising . . . 14

3.1.1 Comparison with related models . . . 16

3.2 Comparative Statics: Blur effect of advertising . . . 18

3.2.1 Price elasticities and blur parameter . . . 21

3.2.2 Relationship of the blur parameter and single product advertising . . . 23

4 Identification of Nested Consideration Logit Model **25**

4.1 Identification with higher level nests . . . 27

4.2 Discussion about **IA2** . . . 28

4.3 Parametric identification of blur parameter . . . 29

5 Conclusion **30**

A Appendix: Main Derivations **31**

A.1 Derivation of Nested Consideration Logit model . . . 31

A.2 Derivatives for attention effects . . . 34

1 Introduction

Understanding how advertising shapes consumer behavior is a longstanding concern in both economics and marketing. Traditional models often treat advertising as either enhancing product image (prestige effect) or improving consumer information (information effect). However, these two roles are often entangled, and existing frameworks struggle to capture their distinct impacts, especially when market level data is the only data allowed.

This paper introduces a structural model—the Nested Consideration Logit (NCL) model—that explicitly separates and analyzes two central advertising effects: the *attention effect*[1], which increases the likelihood that a product enters a consumer's choice set, and the *blur effect*, which diminishes the consumer's ability to perceive differences in quality across products. NCL can be practically used in applied works in analyzing advertisements when only aggregate level data is available. Our model incorporates limited consideration problems and information effects (blur) of advertising while also allowing for the *spill-over effect* of advertising on other products in the same industry. The attention effect and blur effect parameters, along with the other model parameters, are identified with the existence of IV and certain data requirements.

The NCL model builds on the nested logit framework but incorporates consideration probabilities and information distortions in a novel way. Specifically, it allows advertising to influence both the probability of industry selection (nest-level) and the product choice within the industry (product-level)[2]. A key innovation is the modeling of the blur effect as a scalar parameter that reflects the aggregate advertising intensity within a nest, reducing the salience of product-level attributes and pushing consumers to rely more heavily on relative advertising-driven impressions (**Fact 1**). The attention effect, by contrast, is captured through product-specific advertising that raises consideration probabilities directly.

The inclusion of spill-over effects, where advertising for one product can affect consideration probabilities for other products in the same industry, naturally arises by modeling choice into a stage-wise behavior. This creates strategic interactions among firms: in some settings, advertising can behave like a public good, benefiting all firms in a nest by drawing attention to the category (potentially incentivizing free-riding), while in other contexts it becomes a public bad, intensifying competition and diluting individual firm returns. The model thus provides a flexible framework to study the dual nature of advertising competition and its implications for firm strategy.

Comparative statics reveal rich ground for strategic behavior from the model. Firms with lower-than-

[1] In this paper, attention and consideration are regarded as the same object, which is a limited resource of consumer, and for which advertising is trying to seek. We also use 'image' in a similar meaning with attention and consideration.

[2] Coming up with a nested consideration model, where both nest choice and product choice stages are affected by the consideration probabilities of product and nest, respectively, is not very new (Pancras [2011][24], for example).

average perceived product quality may benefit from blurring, as it reduces differentiation, while high-quality firms may suffer. The nest-level demand may also fall if the industry-level advertising makes too much noise. Price elasticities also shift in nontrivial ways: blur generally reduces elasticities (making demand less sensitive to price), both own and cross, but there is an additional term that varies by a price-increasing firm's relative position in the market. The fact that blur benefits lower-utility products and the benefit from blur is shared through price elasticity implies potential possibility of reduced price competition. Furthermore, the fact that the blur parameter will be determined jointly and the effect from it will be heterogeneous suggests how firms will exploit advertising not only as a tool to increase their own demand through raising overall attention. Giving the demand-side grounds for various strategic behavior to rise will be our contribution.

From an empirical standpoint, the introduction of blur and attention effects introduces new identification challenges. Identifying the blur parameter raises similar issues as identifying the nest or scale parameters, but makes the task more difficult by introducing the additional challenge of distinguishing blur from nest parameter. Our strategy here is to exploit a panel data structure. Assuming a data-generating process of unobserved utility and the existence of advertising that gives the same blur parameter for each nest, we can nonparametrically identify the blur parameter along with all other parameters with the minimal normalization of a parameter. Although the first assumption is a restriction, it is a standard BLP assumption. The second assumption may be stronger, but when there are situations where advertising expenditures have not changed a lot for two consecutive periods or when only the composition of advertising firms has changed, not the target values, then this will be satisfied. The assumptions will give us a nonparametric identification of the blur parameter (**Proposition 1**). Also, if the blur parameter is a function of variables in addition to advertising, then with an even stronger assumption the identification follows (**Corollary 1**).

For parametric identification, the blur parameter will be some function of the salience index (HHI polarization index among the nest advertising) and the 2nd-order polynomial of the total advertising of the nest to incorporate the information capacity of consumers in our model. We also consider a market-level data proxy of the learning process as a function of previous market share vectors entering the blur parameter. These parameters are identified based on the nonparametric identification established earlier (**Lemma 1**).

There are several limitations to our model. First, it does not account for differences in the quality of advertising. In practice, firms that recognize both the attention-grabbing and informational roles of advertising may strategically design higher-quality, more informative ads to enhance the information effect, or the reverse, deliberately making more noise. As a result, the effectiveness of advertising may vary even when its measured intensity remains the same. Our model, however, treats advertising intensity uniformly and cannot capture such variation in ad content or quality.

A second, and perhaps more significant, practical limitation concerns the assumption of an exogenous nest structure. Our framework requires the researcher to define product nests in advance, which can be particularly restrictive in advertising contexts. Even when the chosen nesting structure seems reasonable, advertising spillovers may not conform neatly to these partitions. Consumers exposed to an advertisement may form associations with products across multiple nests, not just within the advertised category. While the model could, in principle, be extended to accommodate overlapping nests, the fundamental issue of relying on an exogenously specified spillover structure would remain. Therefore, the model is best suited to settings where the researcher can credibly assume alignment between the nest structure and the scope of advertising spillovers.

Lastly, some concerns stem from the model's identification requirements, particularly the assumption that the blur parameter remains fixed over time. Consistent estimation relies on having multiple time periods during which the relevant aggregate advertising levels are stable. The more such periods exist, the more reliably the model can be estimated. As a result, the model benefits from longer time-series data.

Additionally, for the first-stage parameters to be properly identified, the number of products within each nest must not be too small. This creates a practical tension: although the researcher is free to define the nest structure, there is a natural resistance to forming highly granular nests, as doing so may undermine identification by reducing within-nest variation[3]. Thus, while the nest structure is under the researcher's control, its design must balance conceptual clarity with empirical feasibility.

This paper is organized as follows: In section 2, we review the relevant literature on the topic. In section 3, we introduce the nested consideration logit model and derive its main comparative statics, specifically the effect of advertising through attention and the blur effect on choice probability. In section 4, we discuss various identification results of the nested consideration logit model. Finally, we conclude in section 5.

2 Literature Review

As far as the author knows, the categorization of attention effect and blur effect has not been shown up in the advertising literature. Instead, the categorization of image (or prestige) effect and information effect are used more often (Bagwell [2007][5]). Attention and blur effects can also be considered as image/prestige and information effect, respectively, in our model. Although the way we adopt the core is rather tangential, the distinction is more motivated by Johnson and Myatt [2006][17], where they distinguish the two effects of advertising as "pure hype" and "real information." Although the terminology sounds rather banal, its

[3] Note that we cannot rely on time variation for this because the first-stage parameters are going to be time-dependent by the existence of the time-varying blur parameter.

implication on consumers' willingness to pay distribution is attractive. Hype is related to the mere *shift* of the distribution, whereas giving the information is associated with inducing *dispersion* to the distribution. The shift is translated to the demand curve shift, while the dispersion is again translated to the *rotation* of the demand curve.

In our model, attention effect is a shift made by the advertising as is in Johnson and Myatt [2006]. Blur effect matches with the rotation defined in their paper with a rotation point of a deflated average utility within the nest for the demand function (the choice probability) of a product. Moreover, in Johnson and Myatt [2006], more dispersion is related to more information whereas in our model, blur, which leads to less information, is not related to the dispersion. Dispersion will be determined by the nest parameter.

Although we use different terminology, we contribute on strategies to disentangle the two significant effects of advertising. Trial of distinguishing the image (or prestige) effect and information effect of advertising has been a long-standing issue in the advertisement and recommendation literature both in economics and marketing (Ackerberg [2001,2003][1, 2], Honka et al. [2017] [16], Kawaguchi et al. [2021][18], Cheung and Masatulioglu [2024][11], etc.).

The distinction between these two effects is important because they have different implications for both consumer's welfare and firms' advertising strategies. Preference changes coming from image/prestige effect may be treated as distortion of consumer welfare and could be subject to a control for counterfactual analysis. On the other hand, information effects are more inclined to be depicted as having virtuous role, making people feel less uncertain about product quality. For the supply side, firm's strategy will differ when the efficacy of ads are more of an informative effect than image/prestige effect since compared to the latter effect, the former does not guarantee that the firm benefits from it. Consumers may become sensitive to the difference between the qualities of the product and end up not choosing the product that was advertised.

One of the earliest trial of disentanglement of the two effects was Ackerberg [2001,2003][1, 2]. Attention effect in Ackerberg [2001][1] is defined as the effect of advertising to purchase of consumers who never experienced a newly introduced product, whereas image/prestige (utility) effect is the effect of advertising to both group of consumers who have experienced and not experienced the product. By making use of the entry of new product (Yoplait 150) Ackerberg [2001][1] shows that information effect is prominent, while the utility effect does not seem to be effective. Ackerberg [2003][2] forms a structural model with the learning process of individual consumers and call the effect of advertising that affects through learning process[4] as information effect and the part that directly enter the utility as utility effect with the similar conclusion

[4] Erdem and Keane [1996][12] has the same specification of advertising which enters through Bayesian learning process, but Ackerberg [2003][2] also adds advertising intensity in utility term to distinguish two effects separately.

with Ackerberg [2001][1].

More recently, Honka et al. [2017][16] obtained a rich consumer-level survey data set of bank account choices and takes more granular approach to consumer decision making: awareness, consideration sets, and choice[5]. The results were rather similar with the papers above: ads heightening knowledge (awareness) but not affecting the choice itself that much. Unlike these papers, we are dealing with market level data which does not allow some observations individual level data can reveal. Moreover, our paper allows spill-over to be involved in play, the part these papers have not accounted for.

The possible spill-over may be critical in understanding the effect of advertising since ignoring the effect may bias the effects of advertising. For example, consider a two period panel data of sales and advertising. If, for the first period, the spill-over creates public goods game and public bads game in the second period, the data will show joint increase of advertising. Still, there may not be an increase in sales since consumers' attentions and budgets are limited, so that the joint increase does not change the share of attention. In this case, naive approach might end up concluding advertising does not have a big effect. However, the real effect of advertising will be revealed only when we see a deviation from the equilibrium strategy. The story here tells us how ignoring the spill-over of advertising can be misleading.

Although the terminology used is "recommendation", Kawaguchi et al. [2021][18] and Cheung and Masatulioglu [2024][11] try to disentangle between attention channel (whether a product is considered, information effect) and utility channel (how consumers are affected by the recommendations) allowing the spill-over effects to occur. Both of them formulates consideration set model along with utility function where the effect of recommendation enters to not only to consideration but also the utility. However, the former is an empirical work, while the latter focuses on theoretical characterization. Kawaguchi et al. [2021][18] use a detailed transaction data from vending machines and takes the real-time interface shown on vending machines as recommendations, essentially the same as advertising. Menu variables like number of total or recommended products are only entering to attention to separate the effect of recommendation on utility. They show that, unlike the results above, both attention and utility increases by recommendation. Additionally, worth to note is that in p.5652, they show that the number of recommendation has negative coefficient for the squared term suggesting that too much recommendation has negative effect on attention. This aligns with our intention of tuning the model with the blur parameter but the effect will control for the difference between product qualities, not attention.

Cheung and Masatulioglu [2024][11] shows abundance of behavioral phenomena that occurs through

[5] Here, the awareness means more of a knowledge of the existence of a particular bank, whereas the consideration means more of a search and investigation, a serious knowledge useful for choosing the bank.

recommendation and gives characterization that can disentangle the two: proportional gain (the ratio of choice probability of a product when it is recommended to not being recommended should be larger if the total menu is larger) in choice probability is attributed to utility effect while the sales effect, an increase of total purchase probability, is to attention effect. Although they give insights to how we perceive the two effect differently, their effect of recommendation gives binary recommendation status (recommended or not) rather than grade, which needs a revision on the model to use it in an advertising intensity data.

Moreover, the spill-over in Cheung and Masatulioglu [2024][11] arise through Manzini and Mariotti [2014][22]'s random consideration set mechanism. There, each product has a unique consideration parameter (probability) that is independent of others but constructs a consideration set with a probability built by a multiplication of those parameters of products inside the set. Our model, in contrast, make use of nest structure where positive spill-over only arise for the advertising of rival products in the same nest (industry).

Until now, the papers we have gone through set the attention effect as an information effect contrast to our specification. In our model, although attention is entering the utility function directly, still it has some effect of information if the information is mere awareness, since advertising more will increase the utility. Therefore, instead of using the terminology of information effect and image/utility effect, we stick into attention effect and blur effect. As mentioned above, *shift* and *rotation* of Johnson and Myatt [2006][17] is more closer to what we mean with each advertising effects.

Shapiro [2018][25] has directly dealt with the spill-over issue with the choice of pharmaceutical (antidepressants). In his reduced form approach, he comes up with a "borderline" strategy which exploits similar traits two counties that share a border to control for several unobserved fixed effects. The result shows a significant positive coefficient for the rival's advertising that suggest the existence of positive spillover in advertising. Moreover, in his structural model of two-level independent logit (independent logit for each stage), approves this once more and shows that when the benefits are internalized, firms advertise more, hinting the existence of free-riding in the anti-depressant market. Our model has a similar form with Shapiro [2018][25]'s structural model but has even more structure and we also focus on the blur effect that is out of Shapiro [2018][25]'s scope. Also, his paper is not interested in disentangling the two effects of advertising but rather choose to use advertising as entering the utility function.

Model-wise, the Nested Consideration Logit model is not a very genuine model in terms of having consideration probability playing a role of adjustment factor in front of desirability part of product and nest, both. In Pancras [2011][24], nested consideration model, which is independent logit for each choice stage, has been introduced with consideration probability for each stage. Also, spill-over of advertising can be rather simply achieved through modelling choice behavior into stage-wise form. For example, Shapiro

[2018][25] models the choice of antidepressants as a three-stage process where for each stage stochastic choice follows logit independent of the other stages and allows the advertising stock of each level to affect the choice probability. Unlike these models, however, our model is different in a way that it is micro-founded with the utility maximizing behavior of consumers who care for both product image and industry image formed by ads. This gives us certain structure between consideration probability of products within a nest and the nest itself.

NCL is equivalent to the nested rational inattention model in terms of functional form of choice probability (Fosgerau et al. [2020][13]), instead of blur parameter that divides the systematic utility. However, the consideration probability of our model is not a prior of a consumer that should necessarily match with the mean of posteriors, that is, this is not a rational inattention model. The probabilities are just formed by the advertising competition of firms representing how much consideration they can grab from consumers' limited attention. Moreover, our nested consideration logit model captures aggregate behavior rather than individual consumer's behavior who has nest parameter as one's information cost. In our model, blur parameter is more of information cost. In contrast, nest parameter just controls the dispersion of within nest preference heterogeneity. The Definition 7 of "Generalized nested logit" model in Kovach and Tserenjigmid [2022][20] makes our model a special case of it since we do not allow overlapping nest. However, their focus is to characterize the nested logit and a more broad concept Nested Stochastic Choice (NSC) with behavioral axioms. We focus more on what the specific form of nested logit (NCL) implies for consumer behavior and whether model parameters can be identified.

Related with the equivalence of rational inattention and additive random utility model (ARUM) discussed in Fosgerau et al. [2020][13], derivation of NCL choice probability from our model depends on the perturbed utility representation of the additive random utility model discussed in Allen and Rehbeck [2019][3], Example 1. Moreover, the way the consideration probability should enter into utility is more tractable when working with the perturbed utility model than in additive random utility model where we have to dissect probability distribution to understand the grouping structure. For example, from Allen and Rehbeck [2023][4] where we can find the cost function of nested logit:

$$-\sum_{n\in\mathcal{N}}\left[\eta_n\sum_{j\in n}p_j\ln p_j+(1-\eta_n)\left(\sum_{j\in n}p_j\ln\left(\sum_{j\in n}p_n\right)\right)\right],$$

where $\mathcal{N}$ is the partition of total choice set $\mathcal{J}$, p_j is choice probability of j and η_n being nest parameter n. We were able to find out that the consideration probability of product γ_j and nest $\gamma_n=\sum_{j\in n}\gamma_j$ should enter into systematic utility as $\eta_n\ln\gamma_j+(1-\eta_n)\ln\gamma_n$.

The term “consideration probability” in our model is slightly different from “consideration set probability” that are frequently being used in theory and applied works. For example, alternative specific consideration probability that forms consideration set has been introduced in Manski [1977][21] and formalized in theory in Manzini and Mariotti [2014][22]. In our model, consideration probability does not directly determine the probability of the product being inside the consideration set or not. The same process is rather done by nested logit's first stage decision, which can be understood as menu choice stage. The menus are partition of the total choice set which can be thought of as Luce consideration set model dealt in Brady and Rehbeck [2016][7]. Our consideration probability will be controlling for how intensive the presence of a product will be in attention, but not forming the consideration set itself to be strict.

Loosely, consideration probability in our model has some similar effect with consideration set probability in terms of approximating the consideration set effects. In marketing literature, Bronnenberg and Vanhonacker [1996][9] discuss the importance of limited choice sets and suggests that logit choice probability of particular product j in limited choice set $\mathcal{L}$, $P_j(\mathcal{L})$, can be written as follows:

$$P_j(\mathcal{L}) = \frac{\mathbf{1}\{j \in \mathcal{L}\}e^{v_j}}{\sum_{j' \in \mathcal{J}} \mathbf{1}\{j' \in \mathcal{L}\}e^{v_{j'}}}.$$

However, since researchers have imperfect knowledge about $\mathcal{L}$, above representation is infeasible. Bronnenberg and Vanhonacker [1996] suggest to use $\pi_j = E_{\mathcal{L}}[\mathbf{1}\{j \in \mathcal{L}\}]$, i.e., probabilities of j being contained in randomly arising $\mathcal{L}$, instead of $\mathbf{1}\{j \in \mathcal{L}\}$:

$$P_j^{\pi} = \frac{\pi_j e^{v_j}}{\sum_{j' \in \mathcal{J}} \pi_{j'} e^{v_{j'}}}.$$

Of course, since the logit probability is not linear, we cannot directly obtain P_j^{π} from expected value of $P_j(\mathcal{L})$ by integrating the occurrence of $\mathcal{L}$. Instead, we can understand this as an approximation, possibly biased, of the consideration set mechanism.

Interestingly, some applications of random consideration set mechanisms conflict with reduced form results. For example, Shapiro et al. [2021][26] isolates quasi-random variation of advertising through controlling for various fixed effects so that the left-over variation of advertising are left with the part that is not co-varying with the demand. From the results that advertising return on investment were close to 0 or negative, they conclude firms are misallocating the resources. In contrast, Goeree [2008][14], by using BLP with product-specific consideration probability, shows that although the advertising elasticity is not large, the negative effect on the other products were larger explaining why firms are advertising. Also, her result illuminates that, when limited consideration is assumed, the markup of the firms are estimated to be 19%, which is much larger than the estimate of 5% from just BLP with full consideration model, emphasizing the usage of limited consideration models. However, unlike Shapiro et al. [2021][26], Goeree [2008][14] does not incorporate spill-over effect of advertising. In our model, we allow for spill-over to occur and also has some property of consideration set models and thus could be a good setting to compare results of Shapiro et al. [2021] with a structural model.

In terms of identification, it is important in nested logit literature that the nest-specific nest parameters are not identified as can be seen in Hensher and Greene [2002][15]. They mention that "normalization invariance" was violated when the nest parameters are allowed to be nest-specific within the context of a two-level partition. This may be coming from non-identification of nest-specific nest parameters. In our model, the nest parameters have more meaning than in the other models since it is a weight consumer puts on the product image whereas one minus the nest parameter is the weight that is attached on the nest image made by the advertising. For this reason, we attempt to identify nest-specific nest parameter and identify them separately.

In our specification, although we focus on one-level partition, unlike Hensher and Greene [2002][15], we can reason for the source of non-identification of nest-specific nest parameter. The non-identification stems from the nonzero unobserved utility term. Typically, in the existence of the nonzero unobserved term in utility for each product, only the difference between these unobserved utilities are identified, which means for each nest there will be two unknowns (location of one unobserved utility and nest parameter) for one equation. This shed light on the challenges we face when we try the same with two-level partition which we will discuss in the identification section.

Moreover, as far as author has investigated, it is not very easy to find the identification of nest-specific nest parameters in the literature. For the nest parameters that are constant across nests, it suffice to normalize the unobserved utility of one product to identify the constant nest parameter. However, to do the same in the nest-specific version, we will have to normalize the location of unobserved utility for *each* of the nest, which is not very favorable for comparison between the inclusive values of nests. The only exception the author found the identification is Kim [2014][19]. The paper identifies nest-specific nest parameter by exploiting zero-valued observed covariates but does not allow unobserved fixed utility term. Here, we give constructive identification of nest parameters that are nest-specific, which can also be a contribution of this paper without normalizing the location of unobserved utilities.

3 Nested Consideration Logit Model

We start with an individual consumer i who chooses one alternative j from a set of alternatives (product space) $\mathcal{J} := \{0, 1, 2, \ldots, J\}$. The products can be categorized into the nests which is a partition of $\mathcal{J}$, where we denote nest space, $\mathcal{N}$, with the representative nest n. The consumer i solves the maximization problem from choosing alternative j:

$$\max_{j \in \mathcal{J}} U_{ij}(\delta_{ij}, \boldsymbol{A}), \text{ where } U_{ij}(\delta_{ij}, \boldsymbol{A}) = \delta_{ij}^{\frac{1}{\phi_n}} (y_i - p_j)^{\alpha} \varphi_{ij}(\boldsymbol{A})^{\eta_n(\boldsymbol{A})} \varphi_{in}(\boldsymbol{A})^{(1-\eta_n(\boldsymbol{A}))}.$$

δ_{ij} is the utility of alternative j for consumer i that comes from the product characteristics, y_i is the income of consumer i, p_j is the price of alternative j, A_j is a measure of advertising intensity of alternative j, $\boldsymbol{A}$ will be its total collection across $\mathcal{J}$, and $\boldsymbol{A}^n$ will denote the collection of A_j where j is in the nest n.

$\varphi_{ij}(\boldsymbol{A})$ is the consideration factor for alternative j, and $\varphi_{in}(\boldsymbol{A})$ is the consideration factor for nest n. The parameter $\phi_n(\boldsymbol{A})$ is the blur parameter that controls for how sensitive consumers responds to the product characteristics within the nest we are considering. For the succinctness, $(\boldsymbol{A})$ may be suppressed. η_n is the nest parameter for nest n. The nest parameter will be assumed to be between 0 and 1 to impose consistency of nested logit choice probability with the utility maximization behavior (McFadden [1978][23]). Following BLP [1995][6], we can model $\delta_{ij} = \delta(\boldsymbol{x}_j, \xi_j; \beta) = \exp(\boldsymbol{x}_j\beta + \xi_j)$, where $\boldsymbol{x}_j \in X \subseteq \mathbb{R}^{|X|}$ for all j, is the vector of product characteristics of alternative j observed to econometrician, while ξ_j is the product characteristic utility component that are not observable to econometrician. β is a vector of the collection of linear parameter β^r for x_j^r for all $r = 1, \ldots, |X|$.

The consideration factors can be decomposed into perception heterogeneity and systematic part again as $\varphi_{ij}(\boldsymbol{A}) = \frac{\gamma_j(\boldsymbol{A})}{e_{ij}\nu_{in}}$ and $\varphi_{in}(\boldsymbol{A}) = \frac{\gamma_n(\boldsymbol{A})}{\nu_{in}}$, where $\gamma_j(\boldsymbol{A})$ and $\gamma_n(\boldsymbol{A})$ are consideration probability of j and n, respectively, and e_{ij} and ν_{in} are the random error terms that capture the heterogeneity in the perception of j and n for i. ν_{in} also enters into product consideration factor, since the nest should be perceived in the first place to perceive the product.

To be specific, for nested logit model to arise, $\{-\ln \nu_{in}\}_n \overset{inid}{\sim} C(\eta_n)$ and $e_{ij} \overset{iid}{\sim} Exp(1)$, where $C(\eta_n)$ is the Cardell's [1997][10] conjugate distribution with nest parameter η_n or scale parameter of $\frac{1}{\eta_n}$. When nest parameter is 1, $C(\eta_n)$ is by definition degenerated at 0, which makes $\nu_{in} = 1$ for all i and the model becomes a logit model, whereas the right tail of the density increases as η_n decreases making ν_{in} distributed more on values larger than 1. Thus, reflecting the fact the mean of e_{ij} is 1, we can understand ν_{in} of nest n to deflate the utility perception with number smaller than 1 in average as dissimilarity decreases. This means that if one nest have larger dissimilarity index than the others, then the products in the nest will seem to be more attractive to more people than the other nests which leads to higher probability of choosing the nest itself. At the same time, the increase in dissimilarity index will increase the importance of product-specific consideration factor while decreasing the importance of nest-specific consideration factor, which is quite intuitive.

The special feature of this model is that the consumer i not only be affected by the consideration probability of the product j itself, but also the consideration probability of the nest n that the product j belongs to. This is a key feature of NCL that allows the consideration probabilities to affect both stage of choices: nest choice stage and product choice stage. This can be interpreted as consumers' behavior of hedging between the image/prestige formed by a particular product Ads and the industry image/prestige formed by the nest-wide advertising.

Now, the choice problem can be transformed into comparison of log utility: $\max_{j\in\mathcal{J}} \ln U_{ij}$, where

$$\ln U_{ij} = \frac{\boldsymbol{x}_j\beta + \xi_j + \alpha\ln(y_i - p_j)}{\phi(\boldsymbol{A}^n)} + \eta_n \ln\gamma_j(\boldsymbol{A}) + (1-\eta_n)\ln\gamma_n(\boldsymbol{A}) - lnv_{in} + \eta_n \ln e_{ij}. \tag{1}$$

Then the difference of log utility between the two alternatives gives arise of nested consideration logit choice probability. Since y_i does not change across alternatives, it is not a restriction to subtract $\alpha\ln y_i$ from log utilities and if given alternatives are rather small price compared to income y_i, $\alpha\ln\left(1-\frac{p_j}{y_i}\right) \approx -\frac{\alpha}{y_i}p_j =: -\alpha_i p_j$. For current specification, since we are going to assume we only have a market level data, let's use α instead of α_i. We will get back to this topic in the random coefficient section. Thus, redefine $v_j := x_j\beta + \xi_j - \alpha p_j$. Then, we get the

$$P_j := \underbrace{\frac{\gamma_{j|n}(\boldsymbol{A})e^{\frac{v_j}{\eta_n\phi(\boldsymbol{A}^n)}}}{\sum_{j'\in n}\left(\gamma_{j|n}(\boldsymbol{A})e^{\frac{v_j}{\eta_n\phi(\boldsymbol{A}^n)}}\right)}}_{=:P_{j|n}} \cdot \underbrace{\frac{\gamma_n(\boldsymbol{A})\left(\sum_{j'\in n}\gamma_{j|n}(\boldsymbol{A})e^{\frac{v_{j'}}{\eta_n\phi(\boldsymbol{A}^n)}}\right)^{\eta_n}}{\sum_{n'\in\mathcal{N}}\left[\gamma_{n'}(\boldsymbol{A})\left(\sum_{j'\in n'}\gamma_{j'|n'}(\boldsymbol{A})e^{\frac{v_{j'}}{\eta_{n'}\phi(\boldsymbol{A}^{n'})}}\right)^{\eta_{n'}(\boldsymbol{A})}\right]}}_{=:P_n},$$

where $\gamma_{j|n}(\boldsymbol{A}) = \frac{\gamma_j(\boldsymbol{A})}{\sum_{j'\in n}\gamma_{j'}(\boldsymbol{A})}$ is the relative attention of j within nest n formed by ads. The derivation of the nested consideration logit choice probability is given in the Appendix A.1 through solving the equivalent problem (perturbed utility specification) although this is not necessary.

We can see from the choice probability that as $\phi(\boldsymbol{A}^n)$ grows larger, the choice within the nest ($P_{j|n}$) depends more on the relative image (consideration probability $\gamma_{j|n}(\boldsymbol{A})$). This justifies the interpretation of trying to blur the information about the nest induces nest parameter to increase (more dispersion of utility and diluting the systematic utility differences) which leads to reputation dependent behavior of people choosing within the nest.

On the extreme, if $\phi(\boldsymbol{A}^n) \to \infty$, then the choice probability within the nest totally depends on relative image/prestige of products, i.e.,

$$P_{j|n} \to \frac{\gamma_{j|n}(\boldsymbol{A})}{\sum_{j' \in n} \gamma_{j'|n}(\boldsymbol{A})} = \gamma_{j|n}(\boldsymbol{A}).$$

This shows that for the competition within the nest, firms with better image/prestige will have more incentive to manipulate $\phi(\boldsymbol{A}^n)$ to be larger. Of course, this may not be true for other firms in the nest, which becomes a source of strategic heterogeneity in choosing advertising intensity. This will be more clear once we see the comparative statics below. Similarly, the choice probability of nest n will converge to nest consideration probability γ_n. If an industry is not large enough, this will make another layer of thoughts about how well advertised the industry is. In the end the choice probability of product j will converge to γ_j, the consideration probability of j within the whole product space.

Fact 1 (Convergence to Relative Image)**.** For each $n \in \mathcal{N}$ and $j \in n$,

$$\lim_{\phi_n \to \infty} P_{j|n} \to \gamma_{j|n}(\boldsymbol{A}), \lim_{\phi_n \to \infty} P_n \to \gamma_n(\boldsymbol{A}), \text{ and } \lim_{\phi_n \to \infty} P_j \to \gamma_j(\boldsymbol{A}) \text{ holds.}$$

3.1 Comparative Statics: Attention effect of advertising

We can derive the comparative statics of the model to see how the change in advertising affects the choice probability. Also, for the separate analysis of the two effects, we will decompose the effect of advertising into attention effect and information effect by assuming individual firms advertising may not change the dispersion parameter. This may be achieved by assuming that the nest parameter is a function of a sufficient statistics, $B_{nt} \equiv B(\boldsymbol{A}_t^n)$, of individual firms' advertising within the nest, instead of directly using the individual firm's advertising.

Assumption MA1 (Sufficiency)**.** $\exists B(\cdot)$ s.t. $\phi(B(\boldsymbol{A})) = \phi(B(\boldsymbol{A}')) = \phi(B)$ for $\forall \boldsymbol{A}, \boldsymbol{A}'$ with $B(\boldsymbol{A}) = B(\boldsymbol{A}') =: B$.

B_{nt} can be thought of as an aggregator that represents *information noise coming from joint advertising* in the nest. If $B(\cdot)$ is a known function, this may allow fixing ϕ while changing individual product's advertising for values of $\boldsymbol{A}^n$[6]. Be mindful that this is not necessary for any of our analysis, but needed just to separately

[6] We will go into details about what happens if a single firm increases their advertising intensity in section 3.2.2. For simple

describe the effect of advertising on attention excluding the blur effect. Moreover, even though the blur parameter must be a scalar, the aggregators that are involved with the blur does not have to be, i.e., B_n can be a vector of possible aggregators constructed from the joint advertising within a nest $\boldsymbol{A}^n$.

Additionally, we are going to specify the consideration probability for product j and nest n as logit form:

$$\gamma_j(\boldsymbol{A}) = \frac{e^{\gamma^n A_j}}{\sum_{k \in \mathcal{J}} e^{\gamma^n A_k}}, \ \& \ \gamma_n(\boldsymbol{A}) = \sum_{j' \in n} \gamma_{j'}(\boldsymbol{A}),$$

where γ^n is the nest-specific parameter that translates advertising intensity into the effect on attention. This specification leads to an intuitive comparative statics that facilitates the understanding of the role of relative reputation along side of relative market share in attention effect.

The effect of advertising on attention is where the spill-over effect arise in advertising. We are going to investigate three cases: 1) when the own product j's advertising increases, 2) when the other product j' in the same nest n increases, and 3) when the product k in the other nest n' increases. It is quite evident that 1) will have a positive effect and 3) will have a negative effect on the choice probability. The interesting part is the effect of 2), whose sign will depend on several environmental factors (relative consideration probability, market share, and nest parameter, etc.). The derivation of the below equations will be given in Appendix A.2.

$$\begin{aligned}
&1)\ \frac{\partial P_j}{\partial A_j} = \gamma^n P_j \left[1 - P_{j|n} + \left\{(1-\eta_n)\,\gamma_{j|n} + \eta_n P_{j|n}\right\}(1 - P_n)\right], \\
&2)\ \frac{\partial P_j}{\partial A_{j'}} = \gamma^n P_{j'} \left[\left\{(1-\eta_n)\,\gamma_{j'|n} + \eta_n P_{j'|n}\right\}(1 - P_n) - P_{j'|n}\right], \\
&3)\ \frac{\partial P_j}{\partial A_k} = -\gamma^{n'} P_j \left[(1-\eta_{n'})\,\gamma_{k|n'} + \eta_{n'} P_{k|n'}\right] P_n.
\end{aligned}$$

The interesting part of this comparative statics is on 2). As already mentioned, 1) is always positive and 2) is always negative. However, the sign of 2) will depend on the relative strength of the two mediating factors: business stealing from the other nests and business stealing from the other nest mates. Observe that the square bracket of 2) is going to appear exactly the same for all the nest mates in the industry (nest). If business stealing from the other nest is larger than the business stealing from the nest mates, then the advertising of the nest mate will be a public goods, which may lead to under-provision of advertising. However, if the business stealing from the nest mates is larger than the business stealing from the other nest, then the advertising of the nest mate will be a public bads. The firms may suffer from an expensive

example, if the increase of advertising intensity for a single product aggravates the information overload but, at the same time, improves the salience of information, this may lead to a fixation of the blur parameter.

equilibrium for competition including advertising when this is the case.

Although the above result is interesting, it is already dealt in Shapiro [2018] also through stage-wise decision making model. Our model differs with the existing models in what the model implies for the business stealing effect of others. Thus, the term $(1-\eta_n)\,\gamma_{j|n}+\eta_n P_{j|n}$ is the key in understanding the difference of our model compared to other models. This can be understood as business stealing power of j in n from all other nests, which will be symmetric across nests by the independence of irrelevant nest (IIN) property of nested logit model. This is why we multiply $1-P_n$ to the factor $(1-\eta_n)\gamma_{j|n}+\eta_n P_{j|n}$, instead of calculating heterogeneous stealing factor for each of the other nest. The effect is the convex combination of the relative reputation of j within the nest ($\gamma_{j|n}$) and the choice probability of j within the nest ($P_{j|n}$). The higher the nest parameter is, the more the choice probability within the nest dominates the relative attention of j within the nest formed by advertising.

Recall that, in (1), consumer hedge the product image with η_n and industry image with $1-\eta_n$. However, for the business stealing from other nest to work, it should go through the nest choice stage which is determined by the inclusive value and the nest consideration probability. The coefficient of the inclusive value which contains all product relevant characteristics including advertisements is η_n, while nest consideration probability's coefficient is $1-\eta_n$. Therefore when there is an increase in advertising of a particular product, the effect of business stealing of other nests will be divided into the part that is being mediated through the nest consideration probability $1-\eta_n$ and the part that is being mediated through the inclusive value η_n.

3.1.1 Comparison with related models

We can compare 2) with other models. We will compare the result with the existing model of Shapiro [2018][25], which can be thought of as independent logit for each stage, and simple nested logit model with advertising intensity entering as a characteristic of a product. To be specific, the indirect utility functions are given as follows given $\mathcal{N}(\{\eta_n\}_n)$ being joint distribution of nested logit with nest parameters $\{\eta_n\}_n$:

$<$ Indirect Utility Comparison $>$

$$\text{Shapiro [2018]}: U_{in} = \gamma_2 \sum_{j \in n} A_j + \boldsymbol{x}_n \beta_2 + \xi_n + \varepsilon_{in}, \qquad \{\varepsilon_{in}\}_n \overset{iid}{\sim} T1EV(1)$$

$$U_{ij} = \gamma_1 A_j + \boldsymbol{x}_j \beta_1 + \xi_j + \varepsilon_{ij}, \qquad \{\varepsilon_{ij}\}_j \overset{iid}{\sim} T1EV(1)$$

$$\text{Nested Logit}: U_{ij} = \gamma^n A_j + \boldsymbol{x}_j \beta + \xi_j + \varepsilon_{ij}, \qquad \{\varepsilon_{ij}\}_j \sim \mathcal{N}(\{\eta_n\}_n)$$

$$\text{Nested Consideration Logit}: U_{ij} = \frac{\boldsymbol{x}_j \beta + \xi_j - \alpha p_j}{\phi(\boldsymbol{A}^n)} + \eta_n \ln \gamma_j(\boldsymbol{A}) + (1-\eta_n) \ln \gamma_n(\boldsymbol{A}) + \varepsilon_{ij}, \qquad \{\varepsilon_{ij}\}_j \sim \mathcal{N}(\{\eta_n\}_n)$$

$<$ Comparison of 2) of the comparative statics $>$

$$\text{Shapiro [2018]}: \frac{\partial P_j}{\partial A_{j'}} = P_j \left[\gamma_2(1-P_n) - \gamma_1 P_{j'|n}\right]$$

$$\text{Nested Logit}: \frac{\partial P_j}{\partial A_{j'}} = P_j \left[\gamma^n P_{j'|n} (1-P_n) - \frac{\gamma^n}{\eta_n} P_{j'|n}\right]$$

$$\text{Nested Consideration Logit}: \frac{\partial P_j}{\partial A_{j'}} = P_{j'} \left[\gamma^n \left\{(1-\eta_n)\, \gamma_{j'|n} + \eta_n P_{j'|n}\right\} (1-P_n) - \gamma^n P_{j'|n}\right].$$

Observe from the comparison above that in Shapiro [2018] [25] model, the effect of business stealing from other nests (γ_2) and from other nest mates (γ_1) are independent since he modeled each choice stage as independent logit. Thus, his model allows each level's effect to be more flexible than the other two models as the effect of the lower level does not need to mediate through an inclusive value to influence other nests. Instead, the nested logit model imposes more structure, as business stealing among nest mates can occur through the lower level's parameters and primitives via the inclusive value. The NCL shares a similar property but also incorporates a direct attention channel—$(1-\eta_n)\gamma_{j'|n}(\boldsymbol{A})$—that enables the relative product image to steal business from the other nests. While Shapiro's [2018][25] model offers greater flexibility, there may be certain patterns it cannot capture that the other two models can. However, if the structural assumptions in those models are misspecified, Shapiro's [2018][25] model is likely to provide a better fit.

In comparing the nested logit and nested consideration logit models, we observe that, because consumer response to the consideration probability of product j within nest n is weighted by η_n, business stealing from other nest mates in the NCL model is not scaled down by the nest parameter, unlike in the nested logit model. Given that both specifications are consistent with utility-maximizing behavior—i.e., $\eta_n \in (0,1)$—the extent of business stealing from within-nest rivals tends to be larger under the nested logit model due to this division by the nest parameter.

Rather, the nested consideration logit (NCL) model allows business stealing from other nest mates to be more pronounced due to its additional structure—a convex combination of relative attention and relative choice probability for product j within nest n. Note that $P_{j'|n}$ already contains $\gamma_{j'|n}$, implying that the NCL model places greater emphasis on the relative consideration a product receives through advertising.

When a product garners substantial attention via advertising, yet its relative market share is not as pronounced as its consideration probability, the NCL model predicts that increased advertising will more likely trigger a public goods-like competitive effect than the nested logit (NL) model. This is because an increase in advertising first boosts $\gamma_{j'|n}$, which in turn raises $P_{j'|n}$. In both models, this increase in $P_{j'|n}$ intensifies business stealing from within the same nest, but in the NL model, the effect is amplified by η_n.

Overall, business stealing across nests is more strongly supported under the NCL model, if the firm who increase advertisement expenditure already commands a significant share of attention within the market. In contrast, in NL that is consistent with utility maximization behavior ($\eta_n \in [0,1)$), the public goods cannot arise at all with non-negative attention effect ($\gamma^n \geq 0$):

$$\gamma^n P_{j'|n}\left[1 - P_n - \frac{1}{\eta_n}\right] < 0.$$

There can be interesting thought experiment that can happen through the comparison above. However, an intensive comparison between the models will be left for future research because it will depend on how we model competition between firms, a supply model. Here, we just provide the intuition that this demand model will facilitate the interesting competition pattern between firms in the market.

3.2 Comparative Statics: Blur effect of advertising

For the blur effect, we will first take a look at the general change of the blur parameter $\phi(B)$ and then scrutinize the effect of the change of advertising of single firm in the different situations. To do this, we will specify what is the possible specification of $\phi(B)$. Let's start with the comparative statics of change in the parameter.

$$\frac{\partial P_j}{\partial \phi(B_n)} = \left(\frac{\partial P_{j|n}}{\partial \phi(B_n)}\right) P_n + P_{j|n}\left(\frac{\partial P_n}{\partial \phi(B_n)}\right).$$

Recall that $P_{j|n} = \gamma_{j|n}(\boldsymbol{A}^n)e^{\frac{v_j}{\eta_n\phi(B_n)}} / \sum_{j'\in n}\gamma_{j'|n}(\boldsymbol{A}^n)e^{\frac{v_j}{\eta_n\phi(B_n)}}$. For the exposition, let $\phi_n := \phi(B_n)$ Then, we can first see the effect of the change in $\phi(B)$ on the choice probability within the nest as follows:

$$
\begin{aligned}
\frac{\partial P_{j|n}}{\partial\phi(B_n)} =& P_{j|n}\left(-\frac{v_j}{\eta_n\phi_n^2}\right) - P_{j|n}\sum_{j'\in n}\frac{\gamma_{j'|n}e^{\frac{v_{j'}}{\eta_n\phi_n}}}{\sum_{j'\in n}\gamma_{j'|n}e^{\frac{v_{j'}}{\eta_n\phi_n}}}\left(-\frac{v_{j'}}{\eta_n\phi_n^2}\right) \\
=& P_{j|n}\frac{1}{\eta_n\phi_n^2}\left[\underbrace{\sum_{j'\in n}\frac{\gamma_{j'|n}e^{\frac{v_{j'}}{\eta_n\phi_n}}}{\sum_{j'\in n}\gamma_{j'|n}e^{\frac{v_{j'}}{\eta_n\phi_n}}}v_{j'}}_{=:\overline{v}_n} - v_j\right],
\end{aligned}
$$

and therefore,

$$
\frac{\partial P_{j|n}}{\partial\phi(B_n)} = \frac{1}{\eta_n\phi_n^2}P_{j|n}\left[\overline{v}_n - v_j\right]\begin{cases} > 0, & \text{if } \overline{v}_n > v_j \\ < 0, & \text{if } \overline{v}_n < v_j \end{cases}.
$$

This shows that the effect of blurring through increasing blur parameter is not always a good strategy for all the firms even within a nest. If the average quality of the nest is higher than the product j, then the firm will enjoy the increase in blur parameter. Intuitively, if people are not very sure about the quality difference, then the heterogeneity across individual with have more power to determine the choice. However, if the product's quality is above the average quality ($\overline{v}_n$), the opposite is true. This depicts how firm will heterogeneously be affected by the blur.

The benefit of blurring, however, does not come without a cost. Now, by taking the derivative of the nest choice probability with respect to ϕ_n, we get the following:

$$
\frac{\partial P_n}{\partial\phi_n} = -\frac{1}{\phi_n^2}\overline{v}_n P_n(1-P_n)\begin{cases} < 0, & \text{if } \overline{v}_n > 0 \\ > 0, & \text{if } \overline{v}_n < 0 \end{cases}.
$$

We can see from above, if the average quality of the nest ($\overline{v}_n$) is higher than 0, the value of outside option, then the increase in blur parameter will decrease the nest choice probability. This is because the increase in blur parameter makes the nest less informative, which leads to less inclusive value of the nest. Given this property, the blur effect can be a double-edged sword for firms as mentioned. It may increase the probability of choosing the product when the quality of it is not above average, but, because of the non-differentiation of quality across products, it may decrease the probability of choosing the industry itself.

It might be less appealing to think that if the overall quality is worse than the outside option, then the increase in blur parameter will increase the nest choice probability. This is coming from the fact that when blur happens, then the average utility does not just decrease, but collapse to 0. However, it could be the case that consumers do not become indifferent between buying or not buying, but just have bad impression on the industry and choose the outside option. If a researcher believes that this should be the case, that is, if it is more probable to believe people should not care for buying a product from the industry at all, we may need to modify the utility function to accommodate this. For example, if we believe people don't get confused about the value of money by buying the product, we may instead write down the utility function as

$$\ln U_{ij}^{*} = \frac{\boldsymbol{x}_j \beta + \xi_j}{\phi(\boldsymbol{A}^n)} - \alpha p_j + \eta_n \ln \gamma_j(\boldsymbol{A}) + (1-\eta_n) \ln \gamma_n(\boldsymbol{A}) - ln\nu_{in} + \ln e_{ij}.$$

Then, by the consequence, the blurring will end up collapsing the utility to $-\alpha p_j$, which will be negative, and the average of these will also be negative. Thus, gaining benefits from nest choice probability will be harder in this case since now the criteria is not the average utility being smaller than 0, the inclusive value of the outside option, but should be smaller than average of utility component coming from price which is naturally negative.

The variation above will give another interesting insight on how the blur parameter will affect the competition within the nest. When the blur parameter is large, and consumers are confident only about prices, they will perceive products as largely homogeneous—differing only in price and advertising intensity. Now, hypothetically, assume the relative consideration probability is also pushed to uniform mass across firms in the industry. Then, the firms should compete on price which pushes the price down, which directly affects their markup. If firms believe that this will be more harmful than suffering from unequal consideration probability, they might maintain status quo of attention shares until an environmental change such as reduce in attention-dominant firm's advertising efficiency or advertising stock decay to occur, so that they can safely surge for relative consideration probability. This result offers a novel implication for firm-level strategic behavior under advertising-driven attention asymmetries. However, again, the analysis should be more rigorous with the supply side modeling, which will be left for future research.

In the end, the choice probability will be affected by the increase in blur parameter through

$$\frac{\partial P_j}{\partial \phi_n} = \frac{1}{\eta_n \phi_n^2} P_j \left[\overline{v}_n - v_j - \eta_n \overline{v}_n (1 - P_n) \right] = \frac{1}{\eta_n \phi_n^2} P_j \left[(1 - \eta_n(1 - P_n)) \overline{v}_n - v_j \right].$$

The term $1 - \eta_n(1 - P_n)$ will be between 0 and 1 if η_n is between 0 and 1. In this case, we can see that getting benefit from the blur effect is even harder. Your product quality should be below a deflated version of your nest's average. This result relates the blur with the *rotation* concept of Johnson and Myatt [2006][17], Definition 1, with rotation point of $(1-\eta_n(1-P_n))\overline{v}_n$ which is a function of rotation parameter ϕ_n and satisfying

$$v_j \begin{Bmatrix} > \\ < \end{Bmatrix} (1-\eta_n(1-P_n))\overline{v}_n \Leftrightarrow \frac{\partial P_j}{\phi_n} \begin{Bmatrix} < \\ > \end{Bmatrix} 0. \tag{2}$$

This differs from a mere *shift* of $\frac{\partial P_j}{\partial A_j} > 0$, i.e., even though individual product's increase (or decrease) of advertising intensity allows increase of the blur parameter, it does not guarantee that the firm will get benefit from it. Also, the same holds when the variation of advertising can decrease the blur. Additionally, since (2) holds for all ϕ_n, we can also say $\{P_j\}$ is ordered by the sequence of rotations as they do in Johnson and Myatt [2006][17].

Following the result, an increase in advertising will be not favorable for the majority of the sellers unless there are outlier products within the industry that boosts up the average product utility so that even the deflated one is larger than the majority of the products within the nest. This highlights an important implication for competition. If the majority of products are not gaining from blur, the joint advertising of the industry may have some momentum that allows firms to behave as if they are coordinating. They might refrain from advertising too recklessly to avoid scenarios such as consumers suffering from information congestion. Also, again, they might suffer from some product's reign in terms of attention since contending them with advertising might raise the blur.

In practice, η_n can be estimated to be larger than 1, which starts to disconnect the nested logit choice probability from utility-maximization behavior in terms of additive random utility model setting that we adopted (see, for example, McFadden [1978][23] for general requirement of GEV distribution consistent with ARUM). However, this is still supported in perturbed utility setting, where complementarity is allowed (see Allen and Rehbeck [2023][4]). In this case, $1-\eta_n(1-P_n)$ may become negative. We can see that this makes more easy for increase in blur to hurt the products within the nest. This suggests that complementarity within a nest may exacerbate the negative effects of information blur on products in the industry.

3.2.1 Price elasticities and blur parameter

When it comes to price elasticity, the nested consideration logit model will have the same price elasticity as the nested logit model except for the existence of the blur parameter. Recall that the upper case Ps are choice probabilities between 0 and 1 and the lower case ps are prices of the subscript product.

$$
\begin{aligned}
\varepsilon_{jj} &:= \frac{\partial P_j}{\partial p_j} \cdot \frac{p_j}{P_j} = -\frac{\alpha}{\eta_n \phi_n} p_j \left[1 - P_{j|n} + \eta_n P_{j|n}(1 - P_n)\right] \\
&= -\frac{\alpha}{\eta_n \phi_n} p_j \left[1 - (1-\eta_n) P_{j|n} - \eta_n P_j\right] < 0 \\
\varepsilon_{jj'} &:= \frac{\partial P_j}{\partial p_{j'}} \cdot \frac{\partial p_{j'}}{P_j} = \frac{\alpha}{\eta_n \phi_n} p_{j'} \left[-(1-\eta_n) P_{j'|n} - \eta_n P_{j'}\right] && \text{for } j, j' \in n \\
\varepsilon_{jk} &:= \frac{\partial P_j}{\partial p_{j'}} \cdot \frac{\partial p_{j'}}{P_j} = \frac{\alpha}{\phi_{n'}} p_k P_k > 0 && \text{for } j \in n, k \in n'.
\end{aligned}
$$

Or rather, we can summarize the all three equations as follows given $j \in n$:

$$
\varepsilon_{jk} = -\frac{\alpha}{\eta_{n'} \phi_{n'}} p_k \left[\mathbf{1}\{j = k\} - (1 - \eta_{n'}) P_{k|n'} \mathbf{1}\{n = n'\} - \eta_n P_k\right] \text{ for } k \in n'
$$

The first inequality for the own price elasticity ε_{jj} can be more easily seen in the first line of the equations. Additionally, the first line gives more intuition on how we should interpret the elasticity. When product j increase the price, then first it will lose its business by being substituted by the nest mates with term $1 - P_{j|n}$ whose coefficient is 1, and also lose its business by the products outside the nest by $\eta_n P_{j|n}(1 - P_n)$ whose coefficient is $\eta_n P_{j|n}$ implying that you would lose more if your nest is rather dispersed (η_n) and your relative market share within the nest ($P_{j|n}$) is larger.

The cross elasticity within the nest is also negative if it is consistent with our utility model ($\eta_n \in (0, 1)$), but may be positive (shows complementarity) if $\eta_n > 1$. The cross elasticity with the product outside of the nest is always positive.

At first glance, the blur might be understood as deflating elasticity to 0, but it is actually not that direct since the blur parameter is also inside the choice probabilities. The effect of blur of the nest (n') whose product (k) is increasing its price on price elasticity will be calculated as follows:

$$
\begin{aligned}
\frac{\partial \varepsilon_{ik}}{\partial \phi_{n'}} &= -\frac{\varepsilon_{jk}}{\phi_{n'}} - \frac{\alpha}{\eta_{n'} \phi_{n'}} p_k \left[-(1 - \eta_{n'}) \frac{\partial P_{k|n'}}{\partial \phi_{n'}} \mathbf{1}\{n = n'\} - \eta_n \frac{\partial P_k}{\partial \phi_{n'}}\right] \\
&= -\frac{\varepsilon_{jk}}{\phi_{n'}} + \frac{\alpha}{\eta_{n'}^2 \phi_{n'}^3} p_k \left[(1 - \eta_{n'}) P_{k|n'} (\overline{v}_{n'} - v_k) \mathbf{1}\{n = n'\} + \eta_{n'} P_k ((1 - \eta_{n'}(1 - P_{n'})) \overline{v}_{n'} - v_k)\right]
\end{aligned}
$$

We can see that, in the first term, the blur parameter make the various price elasticities close to 0, i.e., makes it more inelastic. If the original elasticity was negative, this term will increase the elasticity, and conversely, if the elasticity was positive, it will decrease the elasticity, without changing the sign itself because the effect of term diminishes to 0, when the elasticity itself goes close to 0.

However, we have to be aware of the second term. The square bracket of the second term again depends on the relative position of utility of product k in the nest and also the average valuation of the nest n', $\overline{v}_{n'}$. For $\eta_{n'} \in (0, 1)$, we can see that increase in $\overline{v}_{n'}$ will increase the elasticity. This implies, for own elasticity, that the larger the average utility of the nest, the elasticity will get even more inelastic by the increase of the blur parameter. The reason for this can be explained by the comparative statics section above. It is more easier to get benefit from blur when your product's utility is smaller than the average utility. This makes it intuitively evident that if the utility of the product k is higher than the average of the nest, the blur will make the own elasticity more elastic.

For cross price elasticities, when the price-increasing product is in the nest, a larger nest average utility will make the increase of blur to lead to more elasticity. This may sound unintuitive but makes more sense when we think about how increase of the price of the other nest mate benefit the product j. When the nest mate who were actually at the position of gaining profit from the blur increase the price, the substitution coming from the increased price will take part of the nest mate's sales that was gained from the blur. That is, the benefit from blur is *shared* when the beneficiary decides to increase the price. Of course, when the nest mate was a victim of the blur (say the systematic utility is much higher than the nest average), then the harm will be shared by giving inelasticity, too.

The above applies the same to the situation where product outside of the nest is increasing the price. The only part that changes is that the criterion of the beneficiary or not becomes more challenging since now the requirement of higher than the nest average does not dilutes the requirement of $(1 - \eta_{n'}(1 - P_{n'}))\overline{v}_{n'} - v_k$ being positive.

When putting all together, the blur might reduce the price competition. When certain firm has lower utility, then high blur will make the own elasticity to have two folded effect the direct inelasticity of the first term and push to elasticity of the second term. However, in front of the square bracket of the second term, we see the cubed blur parameter entering the denominator, which means that the push to elasticity may well dampens as the blur gets severe, whereas the first term is only divided by the normal blur term. This lessens the incentive to lower the price.

3.2.2 Relationship of the blur parameter and single product advertising

The caveat of all the comparative statics we have dealt above is that it is not about the change of one specific firms advertising but the parameter that is jointly going to be determined by the firms within the nest. When a single firm advertise, we will model them to consider two ways they contribute to the blur 1) the *Salience* of information coming from ads within the nest and 2) the amount of information (*Information Load*) that the advertising structure of the nest provides (check out 4.3 **Parametric identification of blur parameter** for more concrete example). The way we will model the two mechanically will be dealt in the next section. In this subsection, we will conceptually explain how the increase in advertising of a single firm will affect the each effects.

When the status quo in attention was already oligopolistic, an ad increase of a product with minor share will even out the advertising share of the industry from status quo. Then, consumers will have to spare more perceptual energy on concentrating several different things, which may lead to the decrease of the salience (salience dilution) or suffer from signal clutter. This may lead to an increase of the blur parameter. On the other hand, if a single product captures a disproportionately large share of consumer attention due to its dominant advertising, the increase in advertising of a contender of the product will increase the salience of the information by giving a source of comparison between products. Likewise, increase in advertising of a single product can affect the blur parameter in both ways depending on status quo of relative consideration probability.

From the total amount of information stand point, the increase in advertising can also affect the amount of information in both ways. When there is less joint advertising of the industry, for example, public goods situation, the increase in advertising will increase the total amount of information. However, if the industry is already in a situation where the total amount of advertising is too much that information congestion is happening, the increase in advertising will decrease the total amount of information.

Note that if a product's consideration probability is relatively small, the increase of advertising of the product may not affect the blur parameter a lot so that separate analysis of the two comparative statics will be possible. However, if the product's consideration probability is relatively large, then the increase in advertising of the product will likely affect both the attention and also the blur parameter. For example, if a product's relative consideration probability is already large, the additional advertising of the product will easily boost the blur parameter, making the product the relatively more manipulative. While increasing the relative attention by increasing advertising, the firm may at the same induce a decrease salience and/or increasing information congestion. As we have already seen in the model section, such a change will induce consumers to rely on relative consideration probability. The stark depiction of how dominance in consideration probability can give a product (or a firm) over-powered authority leads us to the potential connection between blur parameter and market structure of a market we are analyzing.

4 Identification of Nested Consideration Logit Model

We can identify this when there is no random coefficient, as can be seen below from individual choice probability.

$$\begin{aligned}\ln\frac{P_{j|n}}{P_{j'|n}} =& \frac{v_j - v_{j'}}{\eta_n\phi(B_n)} + \ln\frac{\gamma_{j|n}(\boldsymbol{A})}{\gamma_{j'|n}(\boldsymbol{A})} \\ =& \frac{\boldsymbol{x}_j\beta - \alpha p_j + \xi_j - \boldsymbol{x}_{j'}\beta + \alpha p_{j'} - \xi_{j'}}{\eta_n\phi(B_n)} + \gamma^n(A_j - A_{j'})\end{aligned}$$

Let $\Delta y_{jj'} := y_j - y_{j'}$ for any variable y. Then, we can rewrite the equation as follows:

$$\ln\frac{P_{j|n}}{P_{j'|n}} = \frac{\Delta x_{jj'}\beta - \alpha\Delta p_{jj'} + \Delta\xi_{jj'}}{\eta_n\phi(B_n)} + \gamma^n\Delta A_{jj'}.$$

This shows that the difference in choice probability between two products within the nest depends on the difference in product characteristics, unobserved product characteristics, and the difference in advertising intensity within the nest. Note that the regression above will rely on characteristic difference across products which means that there *must exist several products within the nest* for the regression to give consistent estimates. By assuming the existence of IV regressors[7], say $\boldsymbol{Z}$, for possible endogeneity of $\Delta p_{jj'}$ and $\Delta A_{jj'}$ ($\mathbb{E}[\boldsymbol{Z}_j \cdot \left(\frac{\Delta\xi_{jj'} - \beta_0}{\eta_n\phi(B_n)}\right)] = \boldsymbol{0}$ for all $j \neq j'$ in nest n) and full rank condition, we can identify the following mixed objects by regressing $\ln\frac{P_{j|n}}{P_{j'|n}}$ for all $j \in n$: $\frac{\beta}{\eta_n\phi(B_n)}$, $\frac{\alpha}{\eta_n\phi(B_n)}$, $\frac{\Delta\xi_{jj'}}{\eta_n\phi(B_n)}$, and γ^n. Observe that $\eta_n\phi_n$ is just a fixed value for a given nest. Also, By doing this for all nest, we obtain these mixed parameters for all n. However, the attention effect parameter γ^n can be separately identified without additional process.

Now, by moving on to the probability of choosing nests, let's make use of the outside option which consists of its own nest. Then, we can identify the nest parameter by regressing $\ln\frac{P_n}{P_0}$ for all $n \in \mathcal{N}$ as follows:

$$\ln\frac{P_n}{P_0} = (1-\eta_n)\ln\left[\sum_{j\in n} e^{\gamma^n A_j}\right] + \eta_n \ln\left[\sum_{j\in n} e^{\gamma^n A_j + \frac{v_j}{\eta_n\phi(B_n)}}\right]$$

$$\underbrace{\ln\frac{P_n}{P_0} - \ln\left[\sum_{j\in n} e^{\gamma^n A_j}\right]}_{\text{Known or identified above}} = \underbrace{\frac{\xi_{j'}}{\phi(B_n)}}_{\text{Unknown}} + \underbrace{\eta_n}_{\text{Unknown}} \underbrace{\ln\left[\sum_{j\in n} e^{x_j\frac{\beta}{\eta_n\phi(B_n)} - \frac{\alpha}{\eta_n\phi(B_n)}p_j + \frac{\Delta\xi_{jj'}}{\eta_n\phi(B_n)}}\right]}_{\text{Identified above}}$$

The denominator term of P_n disappears by subtracting the outside option choice probability. We can see

[7] For IVs for advertising, it would be harder to find advertising shifter that is product specific. Instead we can find several market specific IVs such as the number of local advertising agencies, or the number of local newspapers, etc.

that there are two unknowns for one equation that can arise for each nest. If there are no variation within a nest in this equation in our sample, we cannot learn both the intercept and the slope. This means that we need additional information to identify the nest parameter and blur parameter. To proceed, we introduce a panel structure of our data with time space $\mathcal{T}$. Also, the following two assumptions are needed to identify the parameters of interest.

Assumption IA1 (DGP of Unobserved Utilities)**.** $\xi_{jt} = \xi_j + \varepsilon_{jt}$, where $\{\varepsilon_\tau\}_\tau$ are independently distributed with $E\varepsilon_{j\tau} = 0$ for all $\tau \in \mathcal{T}$.

Assumption IA2 (Existence of Composition-only change)**.** $\forall n \in \mathcal{N}, \exists t, t' \in \mathcal{T}$ s.t. $B = B(\boldsymbol{A}_t^n) = B(\boldsymbol{A}_{t'}^n)$ should hold for some t, t' for all n.

Then, we can get the following proposition.

Proposition 1. Given **Assumption IA1**, **IA2**, and fair amount of products for each nest, if we normalize one element among (α, β), the nest parameter η_n and blur parameter ϕ_{nt} separately are identified, for all $n \in \mathcal{N}$ and for all $t \in \mathcal{T}$ along with the other NCL model parameters.

The proposition only says the identification assumptions above are sufficient condition. There may be an ease of the assumptions possible. We will give a loose proof of **Proposition 1**.

*Proof of **Proposition 1***. Since we assume a panel data, we have the following two equations:

$$\ln \frac{P_{nt}}{P_{0t}} - \ln\left[\sum_{j\in n} e^{\gamma^n A_{jt}}\right] = \frac{\xi_{j'}}{\phi_{nt}} + \eta_n \ln\left[\sum_{j\in n} e^{x_{jt}\frac{\beta}{\eta_n\phi_{nt}} - \frac{\alpha}{\eta_n\phi_{nt}}p_{jt} + \frac{\Delta\xi_{jj't}}{\eta_n\phi_{nt}}}\right] + \frac{\varepsilon_{j't}}{\phi_{nt}}, \tag{3}$$
$$\ln \frac{P_{nt'}}{P_{0t'}} - \ln\left[\sum_{j\in n} e^{\gamma^n A_{jt'}}\right] = \frac{\xi_{j'}}{\phi_{nt'}} + \eta_n \ln\left[\sum_{j\in n} e^{x_{jt'}\frac{\beta}{\eta_n\phi_{nt'}} - \frac{\alpha}{\eta_n\phi_{nt'}}p_{jt'} + \frac{\Delta\xi_{jj't'}}{\eta_n\phi_{nt'}}}\right] + \frac{\varepsilon_{j't'}}{\phi_{nt'}}.$$

The difference between the two equations gives us the following by **IA2**:

$$\begin{aligned}\ln \frac{P_{nt}}{P_{0t}}\frac{P_{0t'}}{P_{nt'}} - \ln\left[\sum_{j\in n} e^{\gamma^n A_{jt}}\right] + \ln\left[\sum_{j\in n} e^{\gamma^n A_{jt'}}\right] &= \frac{\Delta\varepsilon_{jj't}}{\phi_{nt}} + \eta_n \ln\left[\sum_{j\in n} e^{x_{jt}\frac{\beta}{\eta_n\phi_{nt}} - \frac{\alpha}{\eta_n\phi_{nt}}p_{jt} + \frac{\Delta\xi_{jj't}}{\eta_n\phi_{nt}}}\right] \\ &\quad - \eta_n \ln\left[\sum_{j\in n} e^{x_{jt'}\frac{\beta}{\eta_n\phi_{nt'}} - \frac{\alpha}{\eta_n\phi_{nt'}}p_{jt'} + \frac{\Delta\xi_{jj't'}}{\eta_n\phi_{nt'}}}\right].\end{aligned} \tag{4}$$

By **IA1**, we can identify the nest parameter, η_n, through $E[\Delta\varepsilon_{jj't}/\phi_{nt} | \text{identified objects}] = 0$.

So far, we can identify the nest parameter η_n but not the blur parameter ϕ_n. From the ratio of mixed identified objects, we can identify $\frac{\phi_{nt}}{\phi_{nt'}}$ or $\frac{\phi_{nt}}{\phi_{nt'}}$ but not the scale of them. To identify this we will normalize one parameter, say β^1, the first element of β for example, to be 1 or -1. Then, since we can identify $\frac{\beta^1}{\eta_n \phi_{nt}}$, we can back out $\frac{1}{\eta_n \phi_{nt}}$ and then ϕ_{nt} itself. This identification under normalization of scale of β is thoroughly studied within the process of identifying utility index in Allen and Rehbeck [2019][3]. Through this process, we can nonparametrically identify the blur parameter ϕ_{nt}, and then every mixed objects will now be identified separately, which allows us to gain α and β. Then, we can again back out the $\phi_{n\tau}$ for all $\tau \in \mathcal{T}$. $\square$

This shows that if we can assume the DGP for unobserved utilities in **IA1**, then we can identify the nest parameter through $\mathbb{E}[\Delta\varepsilon_{jj't}/\phi_{nt}|\text{identified objects}] = 0$. The identification of the blur parameter will be done by assuming **IA2**, the existence of composition-only change, which requires that the advertising intensities of individual products are allowed to change only in a way the aggregators that are in B_{nt} vector are fixed in the nest. This will hold the blur parameter fixed across time ($\phi_{nt} = \phi_{nt'}$), which pushes (4) to a linear regression without intercept. Note that, for the estimation, the more pair (t, t') that satisfy **IA2**, the more accurate the estimate for η_n will be.

4.1 Identification with higher level nests

We can have a glimpse at the issue of higher level nests. For example, consider the first stage is whether to buy or not to buy at all and from the second stage is the choice of which nest to buy from. Then, now we cannot obtain (3) as we see since there is no outside option that is in the same level as the nest. It is present one level above. Instead we will have to rely on difference between the two nests that are in the same level. Let l denote the higher level nest of 'buying' decision with nest parameter η_l and 0 being not buying. Then,

$$\begin{aligned}\ln \frac{P_{n|l}}{P_{n'|l}} - \ln\left[\sum_{j\in n} e^{\gamma^n A_j}\right] + \ln\left[\sum_{k\in n'} e^{\gamma^{n'} A_k}\right] &= \frac{1}{\eta_l}\left[\eta_n \ln \sum_{j'\in n} e^{\frac{v_{j'}}{\eta_n \phi_n}} - \eta_{n'} \ln \sum_{k'\in n} e^{\frac{v_{k'}}{\eta_n \phi_n}}\right] \\ &= \frac{1}{\eta_l}\left(\frac{\xi_j}{\phi_n} - \frac{\xi_k}{\phi_{n'}}\right) + \frac{\eta_n}{\eta_l} \ln \sum_{j'\in n} e^{\frac{v_{j'}-\xi_j}{\eta_n \phi_n}} - \frac{\eta_{n'}}{\eta_l} \ln \sum_{k'\in n} e^{\frac{v_{k'}-\xi_k}{\eta_n \phi_n}}.\end{aligned}$$

Since the first stage regression of product level is intact by the change we had here, we can still identify the inclusive value less the unobserved utility term be identified. However, now the unknown term is not just $\frac{\xi_j}{\phi_n}$ but $\frac{1}{\eta_l}\left(\frac{\xi_j}{\phi_n} - \frac{\xi_k}{\phi_{n'}}\right)$. This means that even if we identify $\frac{\eta_n}{\eta_l}$ and $\frac{\eta_{n'}}{\eta_l}$ here, we cannot directly back out $\frac{\xi_j}{\phi_n}$ nor $\frac{\xi_k}{\phi_{n'}}$. Also, the requirement of the existence of composition-only change will have to change into another stronger version which needs $\phi_{nt} = \phi_{n't} = \phi_{nt'} = \phi_{n't'}$ should hold. If this is satisfied, we get $\frac{\eta_n}{\eta_l}$ for all $n \in \mathcal{N}$, and $\frac{1}{\eta_l}\left(\frac{\xi_j}{\phi_n} - \frac{\xi_k}{\phi_{n'}}\right)$ for one j for each $n \in l$ fixing k.

Then, from the higher level choice probability, we get the following:

$$\ln \frac{P_{lt}}{P_{0t}} - \ln\left[\sum_{n\in l}\sum_{j\in n} e^{\gamma^n A_{jt}}\right] = \eta_l \ln \sum_{n\in l} e^{\frac{IV_{nt}}{\eta_l}} = \frac{\xi_k}{\phi_{n't}} + \eta_l \ln \sum_{n\in l} e^{\frac{IV_{nt}}{\eta_l} - \frac{\xi_k}{\eta_l \phi_{nt}}},$$

where $IV_{nt} := \eta_n \ln \sum_{j' \in n} e^{\frac{v_{j'}}{\eta_n \phi_{nt}}}$. This turn us back to (3) but with higher level choice probability. This shows that by the same argument, we can identify all the parameters of interest, but again with the normalization of one parameter.

4.2 Discussion about IA2

Be aware that to check whether **IA2** holds or not, we have to first specify *which aggregators* are going to be important for the decision of blur parameter. As will be described in the next subsubsection in more detail, several indices that relates human cognition to the joint advertising can be candidates. We can think of a salience measure of information from advertising, the total amount of information measure, etc.

We might think that there are more factors that cause the blur parameter to increase or decrease. For example, consider a relative market shares that varies a lot across time. This may have been generated from different people buying different products in an industry. However, it may also be the case that the same consumers move around different products and learn about their different qualities. Then, they will become resilient to the information blur coming from advertising. This can be thought of as a very sloppy version of learning process. To accommodate the learning process in market level data, we may use variability measure of past market share as a control variable for blur parameter, too. However, this will make **IA2** into another version that is even stronger. Given the essence of the trick in proof to work, now all the variables that is in the ϕ_{nt} should be the same which makes it harder to find such t and t' that satisfies this.

Assumption IA3 (IA2+Additional Variables)**.** $\forall n \in \mathcal{N}$, $\exists t, t' \in \mathcal{T}$ s.t. $B = B(\boldsymbol{A}_t^n) = B(\boldsymbol{A}_{t'}^n)$ and $Z_t^n = Z_{t'}^n$ should hold for some t, t' for all n, where Z_t^n is additional variables that enter the blur parameter, $\phi(B_{nt}, Z_t^n)$.

Then the corollary follows.

Corollary 1. Given **Assumption IA1** and **IA3**, if we normalize one element among (α, β), the nest parameter η_n and blur parameter ϕ_{nt} separately are identified, for all $n \in \mathcal{N}$ and for all $t \in \mathcal{T}$ along with the other NCL model parameters.

Another part to note is that, realistically, one of the most probable situation where **IA2** may occur will be two consecutive periods since firms advertising policy and the actual advertising expenditure may not change in short period of time. However, in those periods, the inclusive value parts may not vary that much since product characteristics will not change that much in a short period of time. This means that the identification will be more precise when the time gap between t and t' is larger. Rather we will have to rely on the variation coming from the prices which are relatively more volatile than the product characteristics.

4.3 Parametric identification of blur parameter

If we rely on parametric formulation of the blur parameter, we can identify not only the blur parameters itself but also how the other variables affect the blur parameter. For the sake of parametric identification, we are going to set two versions of it corresponding to **IA2** and **IA3**, each:

$$\begin{aligned}\phi(\boldsymbol{A}_t^n) =& \psi(1-G_{nt})^\theta e^{-\zeta(A_{nt}-A^*)^2},\\ \phi(\boldsymbol{A}_t^n, D_{nt}) =& \psi(1-G_{nt})^\theta e^{-\zeta(A_{nt}-A^*)^2} D(\boldsymbol{P}_{t-1}^n, \boldsymbol{P}_{t-2}^n)^\mu.\end{aligned}$$

$\boldsymbol{A}_t^n := (A_{j't})_{j'\in n}$ is the advertising intensity vector for nest n, G_{nt} is a measure of salience, or a polarization index of the nest n at time t to be specific, which will be explained below, $A_{nt} := \sum_{j'\in n} A_{j't}$ is the total advertising of the nest n at time t, and A^* is the general capacity of consumers to deal with the total advertising for all industry. The idea is that if the total advertising of the nest is below the threshold, then it doesn't affect the information effect, but if it is above the threshold, then it will affect the information effect. The parameter $\zeta > 0$ controls how fast the information effect decreases as the total advertising deviates from the capacity A^*. θ and ζ is controlling how fast the increase in polarization index and total advertising affects the information effect, each. $\psi > 0$ is a scaling parameter.

For the second specification, we added non-advertising related variable which is two previous periods' choice probability vector of the nest n, $\boldsymbol{P}_{t-1}^n := (P_{j|n,t-1})_{j\in n}$ and $\boldsymbol{P}_{t-2}^n := (P_{j|n,t-2})_{j\in n}$. $D(\cdot,\cdot)$ is a distance measure between the two period. If this distance is large, I regard it as consumers moving around products within the industry, which leads to more resilience to the information blur because of the experience growth through consuming various products. This term can be understood as proxy that controls for consumer learning process with $\mu < 0$, a parameter that controls how much the learning be resilient.

Advertising polarization index for nest n, G_{nt} will be calculated as follows:

$$G_{nt} = 4\sum_{j=1}^{|n|}\left(\frac{A_{jt}}{\sum_{j'\in n} A_{j't}}\right)^2\left(1-\frac{A_{jt}}{\sum_{j'\in n} A_{j't}}\right).$$

This is a adjusted version of HHI (Herfindahl-Hirschman Index) that is widely used in the industrial organization literature to measure market concentration. The polarization index is calculated by multiplying the additional term in the last to penalize the concentration by a single firm. This adjustment is made to ensure that the index ranges from 0 to 1, where 0 indicates perfect polarization (i.e., all firms have equal market share) and 1 indicates perfect concentration (i.e., one firm has all the market share). The number 4 is multiplied in the front to normalize the index between 0 and 1. The highest value ($G_{nt} = 1$) is achieved when advertising share is exactly a half for two firms and 0 for all other firms. The minimum is realized when one firm has all the advertising share, which means no polarization. Also, if $|n| \to \infty$ and each firm has the same advertising share, then the index will approach to 0.

This index is suggested to capture the informativeness of advertising by the salience and the total amount of information that the advertising structure of the nest provides. If a nest has a polarized advertising structure, it will be easier for consumers to distinguish difference between products, which leads to more information effect. On the other hand, if the nest has a more equalized advertising structure, it will be harder for consumers to focus on information among advertising (less saliency), which leads to less information effect. For the total amount of advertising, the more the total advertising is, the more information it provides to the consumers, up to some amount, A^*, which indicates the capacity of consumers to handle information from advertising. This aligns with the information congestion theory by Branco et al. [2016][8]. Empirically,

Of course, since we identified the blur parameters for all n and t under the assumption of **IA1** and **IA2** (or **IA3**), we can just do a linear regression approach to identify the parameters of interest.

Lemma 1. Given **Assumption IA1** and **IA2** (or **IA3**), the following parameters are identify the parameters of interest for the parametric identification of blur parameter, ψ, θ, ζ, and A^* (and μ for **IA3**) with a full rank condition of the relevant variables (G_{nt}, A_{nt}, and A_{nt}^2 (D_{nt} also for **IA3**)).

5 Conclusion

This paper introduces a novel structural framework—the Nested Consideration Logit (NCL) model—that disentangles the dual roles of advertising in shaping consumer demand: the attention effect, which determines whether a product enters a consumer's consideration set, and the blur effect, which reduces perceived differentiation within an industry. By incorporating these effects into a nested discrete choice model with endogenous consideration probabilities, we provide a unified and tractable approach to studying demand in advertising-intensive environments using aggregate data.

In 3 **Nested Consideration Logit Model**, we have shown that consumer depending both product-level attention and the industry image along with additive random utility model will give rise to NCL. In comparative statics in terms of attention effect (3.1), we have seen that the competitor's ad in the same nest will be ambiguous. By comparing this with other models in 3.1.1, we have seen that the ad is more probable to be public goods when the product being advertised has better relative image than the relative market share in NCL.

For the blur effects (3.2), our investigation shows that blur reduces the product differentiation which leads to low-quality products benefit compared to high-quality products. However, nest-level demand may decrease when blur increase depending on average nest utility. The fact that advertise increase attention of a product but may also increase the blur parameter ma give a strategic tension for firms. The price elasticities had general term that gets inelastic by the blur, but still have some room for heterogeneity with respect to the relative position of who are considering to increase price (3.2.1). In the end, it is likely that price competition will decrease if the blur parameter soars since the heterogeneous part dampens faster than the term that drives inelasticity uniformly. In 3.2.2, we built a conceptual framework on how a single product ad can affect the blur parameter in two channel: *salience* and *total information load*.

For empirical practicality, we had to identify for blur parameter along with other model parameters in section 4. The key assumption was time invariance of blur parameter for some times for each nest. The within nest variation will give a mixed parameter and time-variation of inclusive values will separate the nest parameter from the blur parameter by the assumption above. With a minimal normalization all the parameters needed are identified (4.1). We extended the result to higher level nests (4.1) and also discussed the limitations and restrictions of the main assumption in 4.2. Lastly, in 4.3, we identified the parameters that structures the blur parameter.

While the model introduces a flexible and tractable structure for analyzing advertising effects, it also presents limitations. In particular, the assumption of an exogenous nest structure and uniform ad quality may restrict its applicability in more complex or fluid advertising environments. Moreover, the existence of instrumental variable is rather assumed. Nevertheless, the NCL model offers a strong foundation for future empirical work and policy analysis, especially in markets where consumer attention is fragmented, and advertising competition is both strategic and noisy. The possible application to random coefficients nested consideration model and supply side modeling will push the NCL even forward by allowing heterogeneity of perception of blur and rigorously scrutinizing how firms will react to such a demand situation. Additionally, the supply side modeling may give a relationship between advertising decision and other firm-side choice variables which will open the door of control function approach to deal with endogeneity.

A Appendix: Main Derivations

A.1 Derivation of Nested Consideration Logit model

We can derive the model from perturbed utility model with nested logit cost. Since ϕ_n is just a scale factor that is changing v_j, we will ignore the part here and it does not affect the derivation. Also, denote $n(j)$ as the nest j is contained in.

$$\max_{\boldsymbol{p}\in\Delta^{|\mathcal{J}|}} \sum_{j\in\mathcal{J}} p_j \cdot (v_j + \eta_{n(j)} \ln \gamma_j + (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} \gamma_{j'})$$
$$\underbrace{- \sum_{n\in\mathcal{N}} \left[\eta_n \sum_{k\in n} p_k \ln p_k + (1-\eta_n) \sum_{k\in n} p_k \ln \sum_{k\in n} p_k \right]}_{\text{nested logit cost function}}$$

Then, from the fact that $\gamma_j = \frac{e^{\gamma^{n(j)} A_j}}{\sum_{j\in n(j)} e^{\gamma^{n(j)} A_j}}$, we have

$$\begin{aligned}
\eta_{n(j)} \ln \gamma_j + (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} \gamma_{j'} =& \eta_n(\gamma^{n(j)} A_j - \ln \sum_{k\in\mathcal{J}} e^{\gamma^{n(k)} A_k}) \\
&+ (1-\eta_{n(j)})(\ln \sum_{j'\in n(j)} e^{\gamma^{n(j)}} A_{j'} - \ln \sum_{k\in\mathcal{J}} e^{\gamma^{n(k)} A_k}) \\
=& \eta_n \gamma^{n(j)} A_j + (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} e^{\gamma^{n(j)} A_{j'}} - \ln \sum_{k\in\mathcal{J}} e^{\gamma^{n(k)} A_k} \\
=& \eta_{n(j)} \ln \gamma_{j|n(j)} + \ln \gamma_{n(j)}
\end{aligned}$$

The nested logit cost function and perturbed utility representation of nested logit model comes from Allen and Rehbeck [2023][4]. We can see that the attention probability enters into the mean utility as convex combination of the attention of the product and the attention of the nest. Now, by solving this problem through lagrangian we can have:

$$\partial p_j: \quad v_j + \eta_{n(j)} \ln \gamma_j + (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} \gamma_{j'} - \left[\eta_{n(j)} \ln p_j + (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} p_{j'} + 1 \right] - \lambda = 0$$
$$1 + \lambda = v_j + \eta_{n(j)} \ln \gamma_j + (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} \gamma_{j'} - \eta_{n(j)} \ln p_j - (1-\eta_{n(j)}) \ln \sum_{j'\in n(j)} p_{j'} \tag{5}$$

When we denote j' as another alternative that is in the same nest with j, we have

$$\begin{aligned}
& v_j + \eta_{n(j)} \ln \gamma_j - \eta_{n(j)} \ln p_j = v_{j'} + \eta_{n(j)} \ln \gamma_{j'} - \eta_{n(j)} \ln p_{j'} \\
\Rightarrow & p_{j'} = \frac{\gamma_{j'}}{\gamma_j} e^{\frac{v_{j'} - v_j}{\eta_{n(j)}}} p_j \\
\Rightarrow & \sum_{j'\in n(j)} p_{j'} = \frac{p_j}{\gamma_j} e^{-\frac{v_j}{\eta_{n(j)}}} \sum_{j'\in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}}
\end{aligned}$$

Now, by inserting the above to (5),

$$
\begin{aligned}
1+\lambda =& v_j + \eta_{n(j)} \ln \gamma_j + (1-\eta_{n(j)}) \ln \sum_{j' \in n(j)} \gamma_{j'} - \eta_{n(j)} \ln p_j - (1-\eta_{n(j)}) \ln \left[\frac{p_j}{\gamma_j} e^{-\frac{v_j}{\eta_{n(j)}}} \sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}} \right] \\
=& v_j + \eta_{n(j)} \ln \gamma_j + (1-\eta_{n(j)}) \ln \sum_{j' \in n(j)} \gamma_{j'} - \eta_{n(j)} \ln p_j \\
& - (1-\eta_{n(j)}) \left[\ln p_j - \ln \gamma_j - \frac{v_j}{\eta_{n(j)}} + \ln \sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}} \right] \\
=& \frac{v_j}{\eta_{n(j)}} + \ln \gamma_j - \ln p_j + (1-\eta_{n(j)}) \ln \frac{\sum_{j' \in n(j)} \gamma_{j'}}{\sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}}}
\end{aligned}
$$

Now, by denoting k as an alternative either outside or inside of the nest $n(j)$,

$$
\begin{aligned}
& \frac{v_j}{\eta_{n(j)}} + \ln \gamma_j - \ln p_j + (1-\eta_{n(j)}) \ln \frac{\sum_{j' \in n(j)} \gamma_{j'}}{\sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}}} \\
& = \frac{v_k}{\eta_{n(k)}} + \ln \gamma_k - \ln p_k + (1-\eta_{n(k)}) \ln \frac{\sum_{k' \in n(k)} \gamma_{k'}}{\sum_{k' \in n(k)} \gamma_{k'} e^{\frac{v_{k'}}{\eta_{n(k)}}}}
\end{aligned}
$$

$$
\frac{\left(\sum_{j' \in n(j)} \gamma_{j'}\right)^{1-\eta_{n(j)}} \frac{\gamma_j e^{\frac{v_j}{\eta_{n(j)}}}}{\sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}}} \left(\sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}}{\left(\sum_{k' \in n(k)} \gamma_{k'}\right)^{1-\eta_{n(k)}} \frac{\gamma_k e^{\frac{v_k}{\eta_{n(k)}}}}{\sum_{k' \in n(k)} \gamma_{k'} e^{\frac{v_{k'}}{\eta_{n(k)}}}} \left(\sum_{k' \in n(k)} \gamma_{k'} e^{\frac{v_{k'}}{\eta_{n(k)}}}\right)^{\eta_{n(k)}}} = \frac{p_j}{p_k}
$$

$$
\frac{\frac{\gamma_j e^{\frac{v_j}{\eta_{n(j)}}}}{\sum_{j' \in n(j)} \gamma_{j'} e^{\frac{v_{j'}}{\eta_{n(j)}}}} \sum_{j' \in n(j)} \gamma_{j'} \left(\sum_{j' \in n(j)} \frac{\gamma_{j'}}{\sum_{j' \in n(j)} \gamma_{j'}} e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}}{\frac{\gamma_k e^{\frac{v_k}{\eta_{n(k)}}}}{\sum_{k' \in n(k)} \gamma_{k'} e^{\frac{v_{k'}}{\eta_{n(k)}}}} \sum_{k' \in n(k)} \gamma_{k'} \left(\sum_{k' \in n(k)} \frac{\gamma_{k'}}{\sum_{k' \in n(k)} \gamma_{k'}} e^{\frac{v_{k'}}{\eta_{n(k)}}}\right)^{\eta_{n(k)}}} = \frac{p_j}{p_k}
$$

Observe that

$$
\begin{aligned}
&\frac{\sum_{j'\in n(j)}\gamma_{j'}\left(\sum_{j'\in n(j)}\frac{\gamma_{j'}}{\sum_{j'\in n(j)}\gamma_{j'}}e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}}{\sum_{k'\in n(k)}\gamma_{k'}\left(\sum_{k'\in n(k)}\frac{\gamma_{k'}}{\sum_{k'\in n(k)}\gamma_{k'}}e^{\frac{v_{k'}}{\eta_{n(k)}}}\right)^{\eta_{n(k)}}}\\
=&\frac{\sum_{j'\in n(j)}\gamma_{j'}\left(\sum_{j'\in n(j)}\frac{\gamma_{j'}}{\sum_{j'\in n(j)}\gamma_{j'}}e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}\Bigg/\sum_{k'\in n(k)}\gamma_{k'}\left(\sum_{k'\in n(k)}\frac{\gamma_{k'}}{\sum_{k'\in n(k)}\gamma_{k'}}e^{\frac{v_{k'}}{\eta_{n(k)}}}\right)^{\eta_{n(k)}}}{\sum_{n\in\mathcal{N}}\sum_{l\in n}\gamma_l\left(\sum_{l\in n}\frac{\gamma_l}{\sum_{l\in n}\gamma_l}e^{\frac{\delta_l}{\eta_n}}\right)^{\eta_n}\Bigg/\sum_{n\in\mathcal{N}}\sum_{l\in n}\gamma_l\left(\sum_{l\in n}\frac{\gamma_l}{\sum_{l\in n}\gamma_l}e^{\frac{\delta_l}{\eta_n}}\right)^{\eta_n}}\\
=&\frac{\sum_{j'\in n(j)}\gamma_{j'}\left(\sum_{j'\in n(j)}\frac{\gamma_{j'}}{\sum_{j'\in n(j)}\gamma_{j'}}e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}}{\sum_{n\in\mathcal{N}}\sum_{l\in n}\gamma_l\left(\sum_{l\in n}\frac{\gamma_l}{\sum_{l\in n}\gamma_l}e^{\frac{\delta_l}{\eta_n}}\right)^{\eta_n}}\cdot\left[\frac{\sum_{k'\in n(k)}\gamma_{k'}\left(\sum_{k'\in n(k)}\frac{\gamma_{k'}}{\sum_{k'\in n(k)}\gamma_{k'}}e^{\frac{v_{k'}}{\eta_{n(k)}}}\right)^{\eta_{n(k)}}}{\sum_{n\in\mathcal{N}}\sum_{l\in n}\gamma_l\left(\sum_{l\in n}\frac{\gamma_l}{\sum_{l\in n}\gamma_l}e^{\frac{\delta_l}{\eta_n}}\right)^{\eta_n}}\right]^{-1}.
\end{aligned}
$$

which is exactly what we used for the nested consideration logit model.

A.2 Derivatives for attention effects

$$
P_j(\boldsymbol{v},\boldsymbol{A})=\frac{e^{\gamma^n A_j}e^{\frac{v_j}{\eta_{n(j)}}}}{\sum_{j'\in n(j)}e^{\gamma^n A_{j'}}e^{\frac{v_j}{\eta_{n(j)}}}}\cdot\frac{\chi_{n(j)}(\boldsymbol{A}^{n(j)})\left(\sum_{j'\in n(j)}e^{\gamma^n A_{j'}}e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}}{\sum_{n\in\mathcal{N}}\left[\chi_n(\boldsymbol{A}^n)\left(\sum_{j'\in n}e^{\gamma^n A_{j'}}e^{\frac{v_{j'}}{\eta_n}}\right)^{\eta_n}\right]},\tag{6}
$$

for each $j\in\mathcal{J}$. $\chi(\boldsymbol{A}^n;\eta_n,\lambda):=\left(\sum_{k\in n}e^{\gamma^n A_k}\right)^{(1-\eta_n)}$. Let

$$
\begin{aligned}
P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A})=&\frac{e^{\gamma^n A_j}e^{\frac{v_j}{\eta_{n(j)}}}}{\sum_{j'\in n(j)}e^{\gamma^n A_{j'}}e^{\frac{v_j}{\eta_{n(j)}}}}\\
=:&\frac{f_j(\boldsymbol{v},\boldsymbol{A})}{\sum_{j\in n(j)}f_j(\boldsymbol{v},\boldsymbol{A})}\\
P_n(\boldsymbol{v},\boldsymbol{A})\qquad=&\frac{\chi_{n(j)}(\boldsymbol{A}^{n(j)})\left(\sum_{j'\in n(j)}e^{\gamma^n A_{j'}}e^{\frac{v_{j'}}{\eta_{n(j)}}}\right)^{\eta_{n(j)}}}{\sum_{n\in\mathcal{N}}\left[\chi_n(\boldsymbol{A}^n)\left(\sum_{j'\in n}e^{\gamma^n A_{j'}}e^{\frac{v_{j'}}{\eta_n}}\right)^{\eta_n}\right]}\\
=&\frac{g_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\sum_{n\in\mathcal{N}}g_n(\boldsymbol{v},\boldsymbol{A})},
\end{aligned}
$$

where $f_j(\boldsymbol{v}, \boldsymbol{A}) = e^{\gamma^n A_j} e^{\frac{v}{\eta_{n(j)}}}$ and $g_n(\boldsymbol{v}, \boldsymbol{A}) = \chi_n(\boldsymbol{A}^n) \left(\sum_{j' \in n(j)} A_{j'}(\boldsymbol{v}, \boldsymbol{A})\right)^{\eta_n}$. Then, for $j' \in n(j)$ and $k \notin n(j)$, note that

$$
\begin{aligned}
\frac{\partial P_j(\boldsymbol{v}, \boldsymbol{A})}{\partial A_j} &= \frac{\partial P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial A_j} P_n(\boldsymbol{v}, \boldsymbol{A}) + \frac{\partial P_{n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial A_j} P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A}) \\
\frac{\partial P_j(\boldsymbol{v}, \boldsymbol{A})}{\partial z_{j'}} &= \frac{\partial P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial z_{j'}} P_n(\boldsymbol{v}, \boldsymbol{A}) + \frac{\partial P_{n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial z_{j'}} P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A}) \\
\frac{\partial P_j(\boldsymbol{v}, \boldsymbol{A})}{\partial A_k} &= \underbrace{\frac{\partial P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial A_k}}_{=0} P_n(\boldsymbol{v}, \boldsymbol{A}) + \frac{\partial P_{n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial A_k} P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A}) \\
&= \frac{\partial P_{n(j)}(\boldsymbol{v}, \boldsymbol{A})}{\partial A_k} P_{j|n(j)}(\boldsymbol{v}, \boldsymbol{A}).
\end{aligned}
$$

Note that $\frac{\partial f_j}{\partial A_j} = \gamma^n e^{\gamma^n A_j} e^{\frac{v}{\eta_{n(j)}}} = \gamma^n f_j$ and $\frac{\partial f_j}{\partial A_k} = 0$ for $k \neq j$.

$$
\begin{aligned}
\frac{\partial P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} =& \frac{\partial f_j(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} \Big/ \sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A}) - f_j(\boldsymbol{v},\boldsymbol{A}) \frac{\partial \sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} \Big/ \left(\sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})\right)^2 \\
=& \gamma^n \underbrace{e^{\gamma^n A_j} e^{\frac{v}{\eta_{n(j)}}}}_{=f_j(\boldsymbol{v},\boldsymbol{A})} \Big/ \sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A}) - f_j(\boldsymbol{v},\boldsymbol{A}) \gamma^n \underbrace{e^{\gamma^n A_j} e^{\frac{v}{\eta_{n(j)}}}}_{=f_j(\boldsymbol{v},\boldsymbol{A})} \Big/ \left(\sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})\right)^2 \\
=& \gamma^n \left(P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A}) - \{P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A})\}^2\right) = \gamma^n P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A}) \left(1 - P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A})\right) \\
\frac{\partial P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial z_{j'}} =& \underbrace{\frac{\partial f_j(\boldsymbol{v},\boldsymbol{A})}{\partial z_{j'}}}_{=0} \Big/ \sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A}) - f_j(\boldsymbol{v},\boldsymbol{A}) \underbrace{\frac{\partial \sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})}{\partial z_{j'}}}_{=\gamma^n A_{j'}(\boldsymbol{v},\boldsymbol{A})} \Big/ \left(\sum_{j\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})\right)^2 \\
=& -\gamma^n P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A}) P_{j'|n(j)}(\boldsymbol{v},\boldsymbol{A}) \\
\frac{\partial P_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} =& \frac{\frac{\partial g_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} \sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A}) - g_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \frac{\partial \sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})}{\partial A_j}}{\left(\sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})\right)^2} \\
\frac{\partial g_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} =& \left(\sum_{j'\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}} \frac{\partial \chi_{n(j)}}{\partial A_j} + \chi_{n(j)} \frac{\partial \left(\sum_{j'\in n(j)} A_{j'}(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}}}{\partial A_j} \\
\frac{\partial \chi_{n(j)}}{\partial A_j} =& \frac{\partial}{\partial A_j} \left(\sum_{k\in n} e^{\gamma^n A_k}\right)^{1-\eta_n} = (1-\eta_{n(j)}) \left(\sum_{k\in n(j)} e^{\gamma^n A_k}\right)^{-\eta_n} \frac{\partial e^{\gamma^n A_j}}{\partial A_j} \\
=& (1-\eta_{n(j)}) \chi_{n(j)}^{-\frac{\eta_{n(j)}}{1-\eta_{n(j)}}} e^{\gamma^n A_j} \gamma^n \\
\frac{\partial \left(\sum_{j'\in n(j)} A_{j'}(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}}}{\partial A_j} =& \eta_{n(j)} \left(\sum_{j'\in n(j)} A_{j'}(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}-1} \frac{\partial f_j(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} \\
=& \eta_{n(j)} \left(\sum_{j'\in n(j)} A_{j'}(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}-1} \gamma^n f_j(\boldsymbol{v},\boldsymbol{A}) \\
\frac{\partial g_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} =& (1-\eta_{n(j)}) \chi_{n(j)}^{-\frac{\eta_{n(j)}}{1-\eta_{n(j)}}} e^{\gamma^n A_j} \gamma^n \left(\sum_{j'\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}} + \\
& \chi_{n(j)} \eta_{n(j)} \left(\sum_{j'\in n(j)} A_{j'}(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}-1} \gamma^n f_j(\boldsymbol{v},\boldsymbol{A}) \\
=& \gamma^n \chi_{n(j)} \left(\sum_{j'\in n(j)} f_j(\boldsymbol{v},\boldsymbol{A})\right)^{\eta_{n(j)}} \left(\left(1-\eta_{n(j)}\right) \frac{e^{\gamma^n A_j}}{\chi_n^{\frac{1}{1-\eta_{n(j)}}}} + \eta_{n(j)} P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A})\right)
\end{aligned}
$$

Observe that $\chi_{n(j)}^{\frac{1}{1-\eta_{n(j)}}} = \sum_{j' \in n(j)} e^{\gamma^n A_{j'}}$ and also $\frac{\partial \sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} = \frac{\partial g_{n(j)}}{\partial A_j}$. Thus,

$$
\begin{aligned}
\frac{\partial g_{n(j)}}{\partial A_j} &= \gamma^n g_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \left(\left(1-\eta_{n(j)}\right) \frac{e^{\gamma^n A_j}}{\sum_{j'\in n(j)} \gamma_{j'}} + \eta_{n(j)} P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A}) \right) \\
\frac{\partial P_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_j} &= \frac{\gamma^n g_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \left(\left(1-\eta_{n(j)}\right) \frac{e^{\gamma^n A_j}}{\sum_{j'\in n(j)} \gamma_{j'}} + \eta_{n(j)} P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A}) \right)}{\sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})} - g_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \frac{\frac{\partial g_{n(j)}}{\partial A_j}}{\left(\sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})\right)^2} \\
&= \gamma^n \left(\left(1-\eta_{n(j)}\right) \frac{e^{\gamma^n A_j}}{\sum_{j'\in n(j)} e^{\gamma^n A_{j'}}} + \eta_{n(j)} P_{j|n(j)}(\boldsymbol{v},\boldsymbol{A}) \right) P_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \left(1 - P_{n(j)}(\boldsymbol{v},\boldsymbol{A})\right) \\
\frac{\partial P_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial z_{j'}} &= \gamma^n \left(\left(1-\eta_{n(j)}\right) \frac{e^{\gamma^n A_{j'}}}{\sum_{k\in n(j)} e^{\gamma^n A_k}} + \eta_{n(j)} P_{j'|n(j)}(\boldsymbol{v},\boldsymbol{A}) \right) P_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \left(1 - P_{n(j)}(\boldsymbol{v},\boldsymbol{A})\right) \\
\frac{\partial P_{n(j)}(\boldsymbol{v},\boldsymbol{A})}{\partial A_k} &= -P_{n(j)}(\boldsymbol{v},\boldsymbol{A}) \frac{\frac{\partial \sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})}{\partial A_k}}{\left(\sum_{n\in\mathcal{N}} g_n(\boldsymbol{v},\boldsymbol{A})\right)} \\
&= -\gamma^n \left(\left(1-\eta_{n(k)}\right) \frac{e^{\gamma^n A_k}}{\sum_{l\in n(k)} \gamma_l(z_l)} + \eta_{n(k)} P_{k|n(k)} \right) P_{n(k)}(\boldsymbol{v},\boldsymbol{A}) P_{n(j)}(\boldsymbol{v},\boldsymbol{A})
\end{aligned}
$$

Finally, omitting the inputs, and let $\frac{e^{\gamma^n A_{j'}}}{\sum_{k\in n(j)} \gamma_k} = \gamma_{j'|n(j)}$, then

$$
\begin{aligned}
\frac{\partial P_j}{\partial A_j} =& \gamma^n P_{j|n(j)}(1-P_{j|n(j)})P_{n(j)} + \gamma^n \left[(1-\eta_{n(j)})\gamma_{j|n(j)} + \eta_{n(j)} P_{j|n(j)} \right] P_{n(j)}(1-P_{n(j)})P_{j|n(j)} \\
=& \gamma^n P_j \left[1 - P_{j|n(j)} + \left\{ (1-\eta_{n(j)})\gamma_{j|n(j)} + \eta_{n(j)} P_{j|n(j)} \right\} (1-P_{n(j)}) \right] \\
\frac{\partial P_j}{\partial z_{j'}} =& -\gamma^n P_{j|n(j)} P_{j'|n(j)} P_{n(j)} + \gamma^n \left[(1-\eta_{n(j)})\gamma_{j'|n(j)} + \eta_{n(j)} P_{j'|n(j)} \right] P_{n(j)}(1-P_{n(j)})P_{j|n(j)} \\
=& \gamma^n P_j \left[\left\{ (1-\eta_{n(j)})\gamma_{j'|n(j)} + \eta_{n(j)} P_{j'|n(j)} \right\} (1-P_{n(j)}) - P_{j'|n(j)} \right] \\
\frac{\partial P_j}{\partial A_k} =& -\gamma^n \left[\left(1-\eta_{n(j)}\right) \gamma_{k|n(k)} + \eta_{n(k)} P_{k|n(k)} \right] P_{n(k)} P_{n(j)} P_{j|n(j)} \\
=& -\gamma^n P_j \left[\left(1-\eta_{n(j)}\right) \gamma_{k|n(k)} + \eta_{n(k)} P_{k|n(k)} \right] P_{n(k)}.
\end{aligned}
$$

References

[1] ACKERBERG, D. A. Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics* (2001), 316–333.

[2] ACKERBERG, D. A. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review 44*, 3 (2003), 1007–1040.

[3] ALLEN, R., AND REHBECK, J. Identification with additively separable heterogeneity. *Econometrica 87*, 3 (2019), 1021–1054.

[4] ALLEN, R., AND REHBECK, J. Revealed stochastic choice with attributes. *Economic Theory 75*, 1 (2023), 91–112.

[5] BAGWELL, K. The economic analysis of advertising. *Handbook of industrial organization 3* (2007), 1701–1844.

[6] BERRY, A., LEVINSOHN, J., AND PAKES, A. Automobile prices in market equilibrium. *Econometrica 63*, 4 (1995), 841–890.

[7] BRADY, R. L., AND REHBECK, J. Menu-dependent stochastic feasibility. *Econometrica 84*, 3 (2016), 1203–1223.

[8] BRANCO, F., SUN, M., AND VILLAS-BOAS, J. M. Too much information? information provision and search costs. *Marketing Science 35*, 4 (2016), 605–618.

[9] BRONNENBERG, B. J., AND VANHONACKER, W. R. Limited choice sets, local price response, and implied measures of price competition. *Journal of Marketing Research 33*, 2 (1996), 163–173.

[10] CARDELL, N. S. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory 13*, 2 (1997), 185–213.

[11] CHEUNG, P., AND MASATLIOGLU, Y. Disentangling attention and utility channels in recommendations. Tech. rep., Working paper, 2024.

[12] ERDEM, T., AND KEANE, M. P. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science 15*, 1 (1996), 1–20.

[13] FOSGERAU, M., MELO, E., DE PALMA, A., AND SHUM, M. Discrete choice and rational inattention: A general equivalence result. *International economic review 61*, 4 (2020), 1569–1589.

[14] GOEREE, M. S. Limited information and advertising in the us personal computer industry. *Econometrica 76*, 5 (2008), 1017–1074.

[15] HENSHER, D. A., AND GREENE, W. H. Specification and estimation of the nested logit model: alternative normalisations. *Transportation Research Part B: Methodological 36*, 1 (2002), 1–17.

[16] HONKA, E., HORTAÇSU, A., AND VITORINO, M. A. Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics 48*, 3 (2017), 611–646.

[17] JOHNSON, J. P., AND MYATT, D. P. On the simple economics of advertising, marketing, and product design. *American Economic Review 96*, 3 (2006), 756–784.

[18] KAWAGUCHI, K., UETAKE, K., AND WATANABE, Y. Designing context-based marketing: Product recommendations under time pressure. *Management Science 67*, 9 (2021), 5642–5659.

[19] KIM, K. I. Identification of the distribution of random coefficients in static and dynamic discrete choice models. *The Korean Economic Review 30*, 2 (2014), 191–216.

[20] KOVACH, M., AND TSERENJIGMID, G. Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy 130*, 9 (2022), 2411–2461.

[21] MANSKI, C. F. The structure of random utility models. *Theory and decision 8*, 3 (1977), 229.

[22] MANZINI, P., AND MARIOTTI, M. Stochastic choice and consideration sets. *Econometrica 82*, 3 (2014), 1153–1176.

[23] MCFADDEN, D. Modelling the choice of residential location. In *Spatial interaction theory and planning*, A. Karlquist, L. Lundquist, and P. Snickars, Eds. North-Holland, 1978, pp. 75–96.

[24] PANCRAS, J. The nested consideration model: Investigating dynamic store consideration sets and store competition. *European Journal of Operational Research 214*, 2 (2011), 340–347.

[25] SHAPIRO, B. T. Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of political economy 126*, 1 (2018), 381–437.

[26] SHAPIRO, B. T., HITSCH, G. J., AND TUCHMAN, A. E. Tv advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica 89*, 4 (2021), 1855–1879.