Internet Advertising with Information Congestion

Dong Ook Choi∗†

April 2014

Abstract

I estimate a structural model of the demand for the display advertising on Korean Internet portals that accounts for the limited attention that website users devote to ad messages. The theoretical model assumes heterogeneity in the attention span distribution of users. I simulate the probability that a user pays attention to a message posted on the website. This, in turn allows me to identify the extent to which the information congestion affects the advertising demand. The estimated model enables me to predict the probability that a message is seen by a user of one particular website in a given month. These probabilities range from .36 to .67 by websites. The probability is decreasing in the intensity of the users’ activity on the website and is increasing in the age and the education level of website’s users. Male users are shown to have lower probability than females do. The results are compared with those given by a model where the information congestion would be ignored. The congestion being ignored, the platform’s market power turns out to be underestimated. I separately estimate the user demand with mixed-logit model and find that 92.17% of users have negative coefficients of the ad level on a platform. Finally, the estimated user-advertising elasticity is shown to be lower than the estimated attention-advertising elasticity.

Keywords: Internet Advertising, Information Congestion, Structural Estimation.

∗Ph.D. Candidate, ESSEC Business School and THEMA, Université de Cergy-Pontoise, France
†This paper is part of my thesis. I specially thank my advisor Régis Renault. This paper has benefited from comments made by participants at seminars and conferences at ESSEC Business School, Telecom Paristech, THEMA – Université de Cergy-Pontoise, EARIE 2013 (Universidade de Évora), and ADRES 2014 (Université Paris Dauphine), and IO workshop at University of Virginia. I thank Simon Anderson, Albert Bemmaor, Federico Ciliberto, Olivier Donni, Nicolas Glady, Nathalie Picard, and Thomas Trégouët for helpful comments and discussions. Financial support from ESSEC Business School and THEMA is gratefully acknowledged.
1 Introduction

The standard model for the advertising-financed media typically considers the network externalities in both sides of its market, treating the viewer’s attention as an unlimited resource. However, in reality the message receiver’s attention is limited. If the quantity of ads on a platform is excessive, it is likely that different ad messages will be crowding out each other. In this paper I study this additional externality in online media markets and show that ignoring this effect can lead to a bias in the demand estimation.

The externality from the limited attention of advertising viewers is often called information congestion.\(^1\) Previous work on this problem usually assumes exogenous ad cost or exogenous receiver’s decision. This study endogenizes both of them by introducing platforms in the model. Ad prices change by the platform’s advertising demand function and the potential consumers choose to participate in a platform depending on the number of ads. With the single-homing participants who have homogenous attention span, a monopoly platform can control the ad level and internalize the congestion effect. I relax the assumption of homogenous consumers and argue that the heterogeneity in the distribution of attention span causes the congestion problem even with a monopoly platform.

As the first attempt to estimate the effect of information congestion, I suggest a structural model of the platform competition given the heterogenous attention span of consumers. Specifically, I look at the display advertising on the Internet websites as an

\(^1\) Consumers’ attention can be seen as a common property to advertisers. Without a platform which rations the quantity of ads, every advertiser who has slightly higher willingness-to-pay than the communication cost would send out a message to consumers. As a result, receivers face a clutter of messages such as flyers in a mailbox or spam e-mails. The problem is that consumers cannot process all the messages they have received due to the limitation in the attention span. This is especially true for the unsolicited advertising. When this is the case, there should be a screening process by the receiver whether it is rational or not. This screening out of messages must be losses for the corresponding advertisers. Therefore, the externality occurs due to the mismatch between the sender and the receiver: the right message for the potential consumer can be crowded out by the other messages or screened out by the receiver.
empirical application. The online media market distinctively features the possibility of the information congestion by the low message delivery cost and easy accessibility. My data of online display ads are fit for this study for three reasons: unsolicited advertising, low business stealing effect (messages are from diversified industries), and no targeting technology.

The structural model is based on two-sided market empirics. I derive the advertising demand following Rysman (2004) in the context of Internet platform competition. I argue that the congestion effect can be identified by incorporating the message processing rates directly with the amount of the network effect from the user side. Therefore, the change in the aggregate ad quantity affects the advertising demand via the scarcity in ad spaces and the scarcity in the attention span. The difficulty comes from the fact that the attention span of a website user is a latent variable that is unobservable to researchers. For this reason, I simulate this variable based on the assumption that the amount of attention can vary by the attributes of each user. In doing so, I also consider the possible endogeneity in formulating attention span with regard to the user’s behavior. This attention model is fitted with my user-level data which includes website access records and demographic information. Also, I separately estimate the user demand for a platform with a discrete choice model which allows a taste variation.

I collect real market data of Internet websites in South Korea from 2004 to 2007. The novelty of this dataset is the perfect match between the two markets (advertisers and users) and the individual choice information in the user market.\(^2\) The websites in this study are the general purpose search engines that provide a variety of contents to users for free and generate revenue by delivering users’ attention to advertisers. Therefore, these websites are good examples of two-sided market. Besides the relatively low advertising cost, the high accessibility for consumers and the fast growing market size make this

\(^2\)The data partly comes from Choi et al. (2012) and I enriched this dataset with the individual-level panel data.
industry appealing to researchers. Another merit comes from the fact that there was a low degree of regulation on Internet advertising at the study period.

As a result, I estimate the probability of users’ processing messages in each website. The range of this probability is shown from .36 to nearly .67, where the median for each website can be as low as .44 and as large as .54. I find that these probabilities vary by websites. The results suggest a possible bias by the traditional model that does not consider the heterogeneity in consumer’s attention. The traditional model predicts lower market power and higher marginal cost than the general model I propose does. Discrepancies in these outcomes imply that the idea of perfect message delivery by platforms can be far from the reality. The result from the attention span equation tells that male, the young, and the less educated have lower attention spans than others. The estimation of mixed-logit in the user demand shows that 92.17% of users feel nuisance with the ads on the websites. The estimation also shows that the attention-advertising elasticity is higher than the user-advertising elasticity on average. Finally, I perform a counterfactual experiment and I find the effect of the ad-avoidance on the ad-levels and the information congestion using the structural model in this paper. These results can shed lights on the policy implications as well as the marketing strategies in the new media industry.

This study contributes to economic theories on information congestion such as Van Zandt (2004), Anderson and de Palma (2009), Anderson et al. (2012), and Anderson and de Palma (2013) by suggesting a model which endogenizes both of the advertising cost and the consumer participation.\(^3\) My basic framework is deeply based on two-sided market literature where network externalities from both sides characterize this type of markets (Rochet and Tirole, 2003; Caillaud and Jullien, 2003; Armstrong, 2006). This study also contributes to the literature on media economics including theories such as

\(^3\)Anderson and de Palma (2012) also studies the information congestion but they focus on the business stealing effect which is not the main concern of this study.
Anderson and Coate (2005) and empirical applications such as Rysman (2004), Kaiser and Wright (2006), and Wilbur (2008). My paper adds to the literature by introducing additional externality of limited consumer’s attention to the demand estimation of media platforms. There are recent studies on search engines and Internet websites including Athey and Gans (2010); Athey et al. (2013); George and Hogendorn (2012); Taylor (2013). My paper differs from these papers in a sense that my study looks at general purpose websites with no targeting and tracking technologies. The effect of information congestion could be more eminent (therefore, more concerned by advertisers) in those platforms with lower level of targeting.

This paper is organized as follows: Theoretical model on the advertising demand with information congestion and on the user demand with a taste variation are presented in Section 2. I describe my dataset and estimation methods in Section 3. Then, I provide the estimation results and the counterfactual experiment in Section 4 and conclude in Section 5.

2 Theoretical Model

2.1 Inverse demand for advertising slots

There are $J$ websites that are indexed by $j = 1, \ldots, J$ in the market and the outside option is indexed by 0. I look at websites that are advertising-financed platforms so they provide contents for free to users and earn revenue by delivering users’ attention to advertisers. Users who access $j$ are exposed to ad messages on that platform. Advertisers purchase ad spaces on $j$ expecting that their messages can reach users on platform $j$. The larger user base a platform has, the more expected profit for the advertisers. This is captured as a positive network effect in the advertising demand function. Facing these two demand
functions, platforms coordinate advertisers and users to be ‘on board.’ These two-sided market interactions may be summarized as follows: when a platform changes the ad quantity, the user demand is affected by that and, in turn, the advertiser demand is shifted by the change in the number of users. I assume also that there is no targeting or tracking technology.\textsuperscript{4} This means that there is no matching between advertisers and consumers through the platforms.\textsuperscript{5} Websites in this model are all independent players competing with each other (no joint ownership structure). I assume the following regarding the behavior of website users:

**Assumption 1.** (single-homing users) *Website users access just one platform in a period.*

Assumption 1 is a traditional discrete choice assumption on viewers in media economics (e.g. Anderson and Coate (2005)). It implies that when sending a message there is no way other than advertising on platform $j$ to reach a user of $j$. This assumption, though restrictive, simplifies the model. The main focus of this study is the effect of the limited attention of consumers. Users are exposed only to the messages of the platform they choose and there is no need to deal with the effect of the second or more impressions which, in turn, affects the behavior of advertisers.

There is no restriction on the platform choice by advertisers. Advertisers can purchase ad slots on multiple websites. In doing so, advertisers need to know how much they can earn from the choice of the website. This depends on how much ad space they buy and how many users (potential consumers) they can reach through the website. I assume

\textsuperscript{4}Targeting and tracking are very interesting issues recently with Internet advertising (Athey and Gans, 2010; Athey et al., 2013). These two technologies are related with the advertising congestion because they can increase match between advertisers and users. However, we found that there was no specific use of these technologies by Internet platforms during our study period (2004–2007) in South Korea. As an empirical application, this could be one of merits in our dataset for investigating congestion problem.

\textsuperscript{5}Targeted advertising/justification in the data/counterfactual in the result. This can complicate the analysis of the effect from the limited attention span.
that homogenous advertisers choose websites based on the expected number of users as well as the ad price. Websites with the largest user base would be preferred by the advertisers but different choices can still be possible by the observable and unobservables characteristics of the website.

In equilibrium, each advertiser sets its marginal revenue from product sales equal to the ad price. I follow Rysman (2004)’s formulation for the advertising demand. The expected exposure for the users on website $j$ is reflected in the look function which I denote as $L_{jt}(a_{ijt}, \phi_{jt})$, where $a_{ijt}$ is the advertiser $i$’s ad quantity and $\phi_{jt}$ is the expected number of views for an ad on platform $j$ in period $t$. Here, $\phi_{jt} = \phi(A_{jt}, U_{jt})$ where $A_{jt}$ is the aggregate number of ads and $U_{jt}$ is the number of user visits to platform $j$ at period $t$. Note that $A_{jt}$ denotes the total number of messages on the platform so that it works to capture the congestion effect in the $\phi$ function. I assume $\frac{\partial L_{jt}}{\partial a_{jt}} > 0$ meaning that the advertiser would get more expected exposure by increasing the amount of its own ad messages, and $\frac{\partial L_{jt}}{\partial \phi_{jt}} > 0$ so that more views on website $j$ increases expected exposures. I impose another assumption to construct a profit function:

**Assumption 2.** (proportionality) *The profit per look of an advertiser is constant.*

Assumptions 1 and 2 imply that an advertiser’s demand for ad space on one platform is independent of how much ad space it purchases on other platforms. This ensures a monopolistic position to each website in the advertising market (“competitive bottleneck” by Armstrong (2006)).

I construct a profit function for the representative advertiser and solve for the profit maximizing ad quantities on all available websites. The total profit of advertiser $i$ from the advertising mix is:

---

6That is, any two look functions for different websites are independent (from the advertiser’s viewpoint): $L_{jt} \perp L_{ij}, \forall j \neq i$. The reasons for this relationship are that advertising on a website targets totally different consumers from the other ones and that the expected profit per look is constant. Thus, an advertiser’s choice of $a$ in one platform does not affect the choice of $a$ in other platforms.
\[ \Pi_{it} = \pi_{it}L_{it}(a_{ijt}, \phi_{jt}) - P_{jt}a_{ijt} + \cdots + \pi_{jt}L_{jt}(a_{ijjt}, \phi_{jjt}) - P_{jt}a_{ijjt}, \]  

(2.1)

where \( \Pi_{it} \) is total profit across all the websites by advertiser \( i \) and \( P_{jt} \) is the ad price on website \( j \). \( \pi_{jt} \) is the advertiser’s average profit per view in website \( j \) and this can be independent from the exposure amount \( L_{jt} \) by Assumption 2. \(^7\) The advertiser decides the number of ads to purchase, \( \{a_{ijt}\}_{j=1}^{J} \), to maximize its expected profit. The price paid for advertising on the website should equal the marginal increase of the expected revenue if advertiser \( i \) puts any ad on platform \( j \). I impose a Cobb-Douglas functional form on \( L_{jt} \) so that \( L_{jt}(a_{ijt}, \phi_{jt}) = a_{ijt}^{\alpha} \phi_{jt}^{\beta} \), where I expect \( \alpha > 0 \) and \( \beta > 0 \) by the assumptions on the look function above. Then, I need the following assumption to model the advertiser’s behavior:

**Assumption 3.** (small advertisers) Each individual advertiser takes the aggregate ad quantity on \( j \) as exogenous. The ad quantity chosen by an individual advertiser has a negligible impact on the total ad quantity of the given website.

This small advertiser assumption allows me to take the aggregate ad quantity \( A_{jt} \), the number of users \( U_{jt} \), and thus, \( \phi_{jt} \) as exogenous to the individual advertiser. Under these assumptions, I solve for \( a_{ijt} \) from the first order condition for the maximization of Eq. (2.1):

\[ \frac{\partial \Pi_{it}}{\partial a_{ijt}} = (\alpha \pi_{jt}) a_{ijt}^{\alpha-1} \phi_{jt}^{\beta} - P_{jt}. \]

There exist corner solutions \( (a_{ijt} = 0, \exists j \in [1, ..., J] \text{ and } \exists t \in [1, ..., T]) \). An advertiser can decide not to choose a website if \( P_{jt} \) is too high or \( \phi_{jt} \) is too low. The optimal ad

\(^7\pi_{ijt} \) for advertiser \( i \) can be used instead of the average profit per view if I allow the heterogenous advertiser’s profit. Rysman (2004) provides related discussion.
level is given as follows:

\[ a_{ijt} = \begin{cases} 
\left( \frac{P_{jt}}{\alpha \pi_{jt} \phi_{jt}} \right)^{\frac{1}{\alpha-1}}, & \text{if the advertiser chooses } j \\
0, & \text{otherwise.}
\end{cases} \]

Let \( \bar{\pi}_{jt} = \pi_{jt} I_{jt}^{1-\alpha} \) where \( I_{jt} \) is the total number of advertisers in website \( j \) at time \( t \). Summing over advertisers, I get \( A_{jt} = I_{jt} a_{jt} = \left( \frac{P_{jt}}{\alpha \bar{\pi}_{jt} \phi_{jt}} \right)^{\frac{1}{\alpha-1}} \). Then, I can derive inverse demand function by solving for \( P_{jt} \),

\[ P_{jt}(A_{jt}, \phi_{jt}) = (\alpha \bar{\pi}_{jt}) A_{jt}^{\alpha-1} \phi_{jt}^\beta. \tag{2.2} \]

The \( A_{jt}^{\alpha-1} \) term in RHS shows [the scarcity effect of the advertising slots]. The decreasing returns in ad spaces on the media \( j \) requires the parameter \( \alpha - 1 < 0 \). \( \phi_{jt}^\beta \) term describes the increasing returns in additional views on the media \( j \) per ad so that the advertiser demand for \( j \) increases in the number of views. Therefore, \( \alpha \) is expected to be less than 1 and \( \beta \) to be positive (these conditions are tested in the estimation). In the next subsection, the formulation of \( \phi_{jt} \) is discussed where the effect of limited attention comes in.

---

8One can think that \( \int_{i \in [0,1]} \left( \frac{1}{\pi_{ijt}} \right)^{\frac{1}{\alpha-1}} dR(i) = \left( \frac{1}{\bar{\pi}_{jt}} \right)^{\frac{1}{\alpha-1}}, \) where \( R(\cdot) \) is a cumulative distribution function of \( \pi_{ijt} \) when heterogenous advertiser’s profits are allowed (see Rysman (2004)).
2.2 User side network effect with information congestion

Information congestion may have different definitions depending on the context. I consider “localized congestion” in the sense that I deal with the effect of information congestion within a certain scope. I first assume that information congestion is confined to one type of media, the Internet. This implies that the effectiveness of Internet ads is independent of other types of media such as televisions, newspapers, magazines, and so on.

I consider display ads on the websites which are an unsolicited way of advertising with no targeting.\(^9\) This implies that advertisers expect to reach average recipients of the population and that users do not know what kinds of messages they will face if they visit a website. The absence of targeting is consistent with general purpose platforms rather than specialized ones. I postulate that the effect of limited attention comes in the expected number of views per ad as \(\phi_{jt} (A_{jt}, U_{jt}) = \phi_{jt} (g(A_{jt}), U_{jt})\), where \(U_{jt}\) is a number of users in website \(j\) and \(g_{jt} = g(A_{jt}) \in [0, 1]\) represents a probability of reaching a user with one ad message.\(^10\)

To generate the probability \(g_{jt}\), I introduce an individual’s attention span, \(m \in \mathbb{R}^+\), which is unobservable to researchers. \(m\) is a random variable representing the maximum amount of messages (ads) processed by a user. The unit of \(m\) is the same as \(A_{jt}\), the number of messages. Let \(F_{jt}(\cdot)\) be the cumulative distribution function of \(m\) on platform \(j\) in time \(t\). Each individual is assigned a value of \(m\) from the distribution (the source of the heterogeneity) and accepts as much as the ratio of \(\min\left(1, \frac{m}{A_{jt}}\right)\) from the messages they have faced. The probability \(g_{jt}\) averages these ratios over all users on website \(j\).

\(^9\)Search engines usually have two types of advertising: display ads and search ads. Display ads in websites are like the ones in newspapers. Users don’t have rights to ask which products would be advertised on the pages. Search ads take an opposite way. In this newer type of advertising, relevant ad messages are shown up when users input keywords of what they want to find.

\(^{10}\)\(g_{jt}\) can be interpreted as the examination function of individual receiver in Anderson and de Palma (2009).
at time $t$ so that $g_{jt} = \int \min(1, \frac{m}{A_{jt}}) \, dF_{jt}(m)$. I show more detailed exposition of $g_{jt}$ and the proof of the differentiability of $g_{jt}$ with respect to $A_{jt}$ in the Appendix. Since I assume that the ad messages are identically and independently distributed among users, the $\phi_{jt}$ function can be generated just by the product of the user visits and the processing rate $g_{jt}U_{jt}$.\footnote{In other words, I assume that there is no externality in the distribution of ad messages and in the user’s examining behavior. The number of views on a website are uniformly distributed within the observation period and ad messages are also distributed in the same way. Therefore, the random variable $m$ is i.i.d. across users. This implies that each visit has the same value to advertisers no matter when the user makes a visit, and thus, each ad message has the same impression value per user in a given observation period.} Therefore, $\phi_{jt}$ is shown as

$$
\phi(A_{jt}, U_{jt}) = U_{jt} \int \min(1, \frac{m}{A_{jt}}) \, dF_{jt}(m).
$$

(2.3)

$\phi_{jt}$ is an expected outcome from the network effect: number of views discounted by the actual ratio of taking ad messages into their cognition process. Allowing the heterogeneity in the distribution of attention span is crucial in my model. If this is not the case, the congestion would be priced out by the platform under Assumptions 1 and 2 with homogenous users as the following proposition shows:

**Proposition 1.** If website users have a homogenous attention span and the marginal contribution of the network effect $\phi_{jt}$ is larger than the marginal contribution of the ad quantity to the platform’s revenue, Information congestion is perfectly priced out. i.e. $A^*_jt \leq m, \forall j$, where $A^*_jt$ is the equilibrium ad level and $m$ is the attention level of homogenous users.

**Proof.** I suppress the subscripts for this proof. Let $m \in \mathbb{R}^+$ be the attention span which is common for homogenous users. Assume that there is information congestion, i.e. $m \leq A$ where $A$ is the aggregate ad level on the platform. From Eq. (2.2),

$$
P = (\alpha \pi)A^{\alpha-1} \left( \frac{m}{A} \right)^\beta U^\beta \text{ where } U \text{ is the total number of user visits and the parameters are}
$$

\footnote{In other words, I assume that there is no externality in the distribution of ad messages and in the user’s examining behavior. The number of views on a website are uniformly distributed within the observation period and ad messages are also distributed in the same way. Therefore, the random variable $m$ is i.i.d. across users. This implies that each visit has the same value to advertisers no matter when the user makes a visit, and thus, each ad message has the same impression value per user in a given observation period.}
$0 < \alpha < 1$ and $\beta > 0$. Therefore, the revenue of the platform is $R(A) = (\alpha \bar{\pi}) A^{\alpha-\beta} m^\beta U^\beta$.

Now if the network effect is larger than the scarcity effect of ad quantity, so $\alpha \leq \beta$, then $R(A)$ is decreasing in $A$. Therefore, the equilibrium ad level should be $A^* \leq m$ where $R(A)$ is the highest, \textit{i.e.} $R'(A^*) = c$. ■

Proposition 1 says that when users are homogenous in the attention span a monopoly platform would choose an equilibrium ad level less than the attention level so that no congestion problem would occur. But, this is for the case when $\alpha \leq \beta$ (the parameter estimates will confirm this condition). When $\alpha > \beta$, the equilibrium ad level can be higher than $m$ because of the weaker contribution to the ad revenue by the network effect. This is the difference between my model and that of Anderson and de Palma (2009). They also gives the same result with the homogenous attention of users but the profit function of the monopoly platform in their model doesn’t account for the size of user base, or equivalently, the nuisance cost in the user demand.

Before moving on, I discuss the difference between my model and Rysman (2004)’s in more detail. First, both models have the structural parameter of the congestion effect but Rysman (2004) cannot identify the parameter separately. When the aggregate ad level rises, the ad price should drop both by decreasing willingness-to-pay for ad slots and by the scarcity of attention. In contrast to Rysman (2004)’s, my model can identify the latter effect by employing the individual-level data on consumer usage. More specifically, the variation of consumers’ responses to the ad level change is the source of identification (for example, the congestion effect can be extremely small with the low level of advertising – because of $\min \left( \frac{m}{A} \right)$ – while Rysman (2004) assumes it to be constant).

Secondly, the interpretation of the congestion effect is different. The information congestion can occur in different stages depending on the types of advertising (see Appendix for the discussion on the various types of advertising). The readers of Yellow Pages are
motivated to search the ads on a specific topic. The competition among the advertisers, therefore, is usually confined to the specific industrial category (e.g. real estates, cars, restaurants, and so on). The congestion happens when the consumers don’t contact advertisers. This is also called “business stealing effect.” Rysman (2004)’s interpretation of information congestion largely depends on this effect. However, this is not the case for my model. I assume that there is no business stealing effect among advertisers (no substitutable ads on one page) because on the general purpose websites like those in my model, users are exposed to the ads from diversified industries rather than from the same industrial category. I assume, therefore, that the only bottleneck is the limited attention span of each individual.

2.3 User demand

Assumption 1 allows me to employ a discrete choice model. I apply a multinomial mixed-logit model in estimating user demand.12 Unlike the simple logit model, mixed-logit doesn’t suffer from the independence of irrelevant alternatives (IIA) property because it allows random taste variation. Users choose a website based on the amount of ads, other exogenous characteristics, and their demographics (in equilibrium users know the total number of ads on all websites in a given period). Price for accessing information or using web services is not charged but users are supposed to watch banners on the page which generates a revenue in the advertising market. In other words, subsidies are given to users because platforms can make profits by selling the attention of users. Advertising is unsolicited. This implies that, ex ante, users don’t know which message they would face until they access the website. Ads take space on the computer screen and distract users from watching contents so the number of ads causes nuisance, which is captured in the user’s utility. This causes a negative network effect in the user demand function. There

12See Train (2003) for more details on the definition and estimation of the mixed-logit model.
is an outside option where users can get information accessing other online sources than the websites in the choice set (or possibly offline services). I begin with the conditional indirect utility function which consists of the website and the user’s characteristics. The utility function of user $k$ when she accesses website $j$ at period $t$ is:

$$u_{jkt} = \rho_k A_{jt} + X_{jt} \lambda_k + c_j^u + \epsilon_{jkt}^u,$$

(2.4)

where $A_{jt}$ is the total number of ads, $X_{jt}$ is a vector of exogenous characteristics of website $j$, and $c_j^u$ is a choice-specific constant. $\rho_k$ is a random coefficient for the ad level with mean $\rho + D_{kt} \omega$ and standard deviation $\sigma^\rho$ where $D_{kt}$ is a vector of demographics of user $k$. $\lambda_k$ is a random coefficient for website characteristics with mean $\lambda$ and standard deviation $\sigma^\lambda$. $\epsilon_{jkt}^u$ is an error term of user $k$ in $j$ and is assumed to follow i.i.d. extreme value distribution. Eq. (2.4) implies that the user’s utility from accessing website $j$ is decided by the amount of ads, exogenous characteristics of the platform, and exogenous shocks. The utility for the outside option (indexed by 0) is specified as $u_{0kt} = \rho_k A_{0t} + X_{0t} \lambda_k + c_0^u + \epsilon_{0kt}^u$, where I set $A_{0t} = 0$ and $X_{0t} = 0$. $c_0^u$ is a constant and $\epsilon_{0kt}^u$ is an extreme value error term for the outside option. Then, the logit probability is derived as $e^{V_{jkt}} \sum_{l} e^{V_{lkt}}$ and so the choice probability of platform $j$ by user $k$ is:

$$P_{jkt} = \int \frac{e^{V_{jkt}}}{\sum_{l} e^{V_{lkt}}} \ dG(\rho, \lambda),$$

(2.5)

where $V_{jkt} = u_{jkt} - c_j^u$ and $G(\cdot)$ denotes a cumulative distribution of random coefficients. A simulation method is used to compute the probability of Eq. (2.5) and the log-likelihood function is derived with this probability. More detailed discussion on the estimation of user demand is provided in the estimation section.
2.4 Equilibrium and efficiency

Assumptions 1 and 2 imply that each platform behaves as a monopolist on the ad market. I assume a constant marginal cost \( c_{jt} \) for website \( j \) at period \( t \). The problem for platform \( j \) at \( t \) is

\[
\max A_{jt} \Pi_{jt}(P_{jt}, \phi_{jt}) = P_{jt}(A_{jt}, \phi_{jt}) - A_{jt}c_{jt}
\]

and so the derivative with respect to \( A_{jt} \) is given as:

\[
\frac{\partial \Pi_{jt}}{\partial A_{jt}} = P_{jt} - c_{jt} + \frac{\partial P_{jt}}{\partial A_{jt}} A_{jt} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial A_{jt}} A_{jt}.
\]  

(2.6)

The third term comes from the direct impact of \( A_{jt} \) on price and the fourth term shows the change in the network effect from the user side. The expected network effect is \( \phi_{jt} = \phi(g(A_{jt}), U(A_{jt}, A_{-jt})) \) so the partial change of \( \phi \) is \( \frac{\partial \phi_{jt}}{\partial A_{jt}} = \frac{\partial \phi_{jt}}{\partial g_{jt}} \frac{\partial g_{jt}}{\partial A_{jt}} + \frac{\partial \phi_{jt}}{\partial U_{jt}} \frac{\partial U_{jt}}{\partial A_{jt}} \).

I plug this into Eq. (2.6), then I can get the equilibrium ad quantity by setting the derivative equals zero as following:

\[
A_{jt}^* = -\frac{P_{jt}^* - c_{jt}}{\frac{\partial P_{jt}}{\partial A_{jt}} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial g_{jt}}{\partial A_{jt}} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial U_{jt}}{\partial A_{jt}}},
\]  

(2.7)

where \( P_{jt}^* \) and other partial derivatives are the realizations at the equilibrium quantity \( A_{jt}^* \). The price-cost margin is given as following:

\[
\frac{P_{jt}^* - c_{jt}}{P_{jt}} = -\left[ \varepsilon_{A_{jt}}^{P_{jt}} g = \bar{g} \text{ and } U = \bar{U} \right] + \varepsilon_{\phi}^{P_{jt}} \varepsilon_{g}^{A_{jt}} + \varepsilon_{\phi}^{P_{jt}} \varepsilon_{U}^{A_{jt}} \right],
\]  

(2.8)

where \( \varepsilon_{A_{jt}}^{P_{jt}} g = \bar{g} \text{ and } U = \bar{U} < 0 \) is the elasticity of the price with respect to the ad quantity under the constant \( g_{jt} \) and \( U_{jt} \), \( \varepsilon_{A_{jt}}^{P_{jt}} = \frac{d \ln P_{jt}}{d \ln A_{jt}} \) is the elasticity of the price with respect to the expected network effect, \( \varepsilon_{A_{jt}}^{g} = \frac{d \ln g_{jt}}{d \ln A_{jt}} < 0 \) is the elasticity of message processing rate with respect to the ad quantity (see the Appendix for more
and $\varepsilon_A = \frac{d \ln U_j}{d \ln A_{jt}} < 0$ is the elasticity of the user demand with respect to the ad quantity. Eq. (2.8) shows that the markup of website $j$ at period $t$ is a combination of these four elasticities and that it must be positive. Roughly speaking, the price-cost margin consists of the traditional market power, the effect of information congestion in the message processing rate, and the effect of the user-side network effect.

The efficient ad level is chosen to maximize the sum of advertiser and platform surpluses, $W = \sum_{j=1}^{J} \left[ \int_{0}^{A_{jt}} P_{jt}(x, \phi(g(x), U_{jt}(x), A_{jt})) - c_{jt} dx \right]$. Then, the first order condition goes:

$$\frac{\partial W}{\partial A_{jt}} = P_{jt} + \int_{0}^{A_{jt}} \frac{\partial}{\partial \phi_{jt}} P_{jt}(x, \phi_{jt}) \frac{\partial}{\partial g_{jt}} g_{jt}(x, U_{jt}(x)) \frac{\partial}{\partial A_{jt}} g_{jt}(x) dx$$

$$+ \sum_{l=1}^{J} \left[ \int_{0}^{A_{lt}} \frac{\partial}{\partial \phi_{lt}} P_{lt}(x, \phi_{lt}) \frac{\partial}{\partial U_{lt}} U_{lt}(x, U_{lt}(x)) \frac{\partial}{\partial A_{jt}} U_{lt}(x) dx \right] - c_{jt}.$$  

Let $A_{jt}^e$ be the efficient level of ads which makes Eq. (2.9) equals to zero. Since websites in the model are effectively being monopolists in the market, the two ad-levels, $A_{jt}^e$ and $A_{jt}$ hardly coincide. Elasticities in equilibrium and the efficient ad level will be predicted with the parameter estimates of the model.

3 Data and Estimation

3.1 Search engine market in South Korea

I look at the Internet search engine market in South Korea considering three agents (the search engines, users, and advertisers) and two markets around the platforms (the user market and the advertising market). I select the six biggest Internet search engines
in Korea: “Naver.com,” “Daum.net,” “Nate.com,” “Yahoo.com/kr,” “Empas.com,” and “Paran.com.” Almost 95% of Internet users in Korea visit one of these websites more than once in a month during the study period. The search engines are content providers in the user market and at the same time they make profits by selling the attention of users in the advertising market. The consumers have no accessibility issues (e.g. geographical restrictions) to the websites and thus, those markets of competing search engines are fully overlapped.

Internet websites (including search engines) are suitable to study information congestion for several reasons. First, they are mostly free and easy to access for consumers so there are more chances to be exposed to ad messages from the consumer’s viewpoint. On average, about half of the Korean population have accessed one of these websites during the study period. Second, the display ads (also called banner ads) in these websites are the unsolicited type of advertising, meaning that the consumers don’t ask for the messages to be shown. Also, there is no targeting or tracking technology used by the websites during the study period. Finally, the cost per impression is lower than that of other types of media. This could encourage advertisers to send messages excessively so it is highly probable for users to suffer from the information overloads.

Yahoo Korea was the first mover in the search engine market in Korea starting its service in 1997. Two years later, Korean-origin services opened in the year of 1999: Naver.com, Daum.net, and Empas.com. Each service has its own strength. Naver.com became the most popular among Koreans after it successfully launched a knowledge sharing service. Daum.net provides a free e-mail service which had the largest subscriber base in Korea. Empas.com is well-known for its unique search engine technology. As these services got popular and the whole market size grew, big telecommunication companies became

---

13 Although Google is the biggest search engine in the world, it had not been successful in some Asian countries such as China and South Korea during the study period. Google’s market share was less than 3% in Korea. Moreover, Google did not provide banner advertising that is the main concern in my study. Therefore, I omitted Google in the study.
interested in this market. In 2001, Nate.com is launched by SK telecom, the largest mobile company in Korea, and Paran.com by KT (Korea Telecom) in 2004.

Naver.com outgrew Yahoo.co.kr in 2003 and Daum.net in 2004. Naver.com has sustained the largest market share after then. Daum.net had been the most popular search engine before 2004 and now it is the second biggest service. One can see the growth of each website in terms of the banner revenue in Figure 1.

3.2 Data

My dataset is composed of two parts: advertising data and user data. Advertising data is at an aggregate level (website/month) and user data is available at an individual level (user/website/month). I discuss each of them in this subsection.

Aggregate level advertising and website information

The advertising data is for online display ads posted on the six search engines. This data is provided by ResearchAd (http://www.researchad.co.kr), an Internet-ads surveying company. The observation period is thirty-three months from June 2004 to February 2007 so there are 198 observations in total (6 websites x 33 months). Descriptive statistics for the advertising data are shown in Table 1.

I have monthly information for each website ad revenue as well as some characteristics I treat as exogenous such as the number of sections and service dummy variables. I also
have the ad-level information and I aggregate them into monthly information.\textsuperscript{14} There is the ad-level price information in the data but I just use one representative price per month per website.\textsuperscript{15} The daily price of the main banner in the top page of each website is used as the representative price for a website in a given month (see Choi et al. (2012)). The ad quantity is generated as a weighted average in a month and the weight is given by the relative daily price of each type of banner.\textsuperscript{16}

The number of sections approximates the amount of content in a website.\textsuperscript{17} This variable counts the number of different sections in the website such as news, blogs, games, and so on. I assume that the decision on contents (and sections) have been made prior to the ad placement decision, so that the number of sections is exogenous to advertising demand.

I use the starting/finishing day of each ad and generate a “contract period” for each individual ad. Then, I aggregate these variables at the website/month level by taking averages. In Table 1, I present average contract periods across websites. Nate.com has the longest average contract period, 7.85 days while Daum.net has the shortest one, 5.9

\begin{equation}
Q_{jt} = \sum_{i=1}^{n} \frac{p_{ijt}}{p_{ijt}} q_{ijt},
\end{equation}

where the subscript $i$ indicates the type of the ad and $n$ is the total number of different types in website $j$ at period $t$. $Q_{jt}$ is the aggregate quantity, $p_{ijt}$ is the daily price of type $i$, and $q_{ijt}$ is the number of ads of type $i$. $\frac{p_{ijt}}{p_{ijt}} q_{ijt}$ means the quantity adjusted by the relative price of type $i$ to the basis price ($p_{ijt}$).

\textsuperscript{14}I have the advertiser’s names, industry classifications, product names, ad contents, contract periods, sizes, and (estimated) prices in ad level. Observations are 38,874 in total during the study period. I set observation intervals as one month in order to match with the data on the user side.\textsuperscript{15}This information is not the real contract price but the estimated one by the surveying company (ResearchAd.co.kr). They estimate the price based on the listed prices, the number of users (the approximation of impressions), ad sizes, and the number of contract days.\textsuperscript{16}First, I count the contract numbers by different sizes of ads in each time period. Then, the price-adjusted quantity in the month is computed as follows:

\textsuperscript{17}There are two different measures for contents amount in previous works. Software variety is used in video games and PDAs market (Nair et al., 2004; Clements and Ohashi, 2005). Actual content size (page number) is used in yellow page and magazine market (Rysman, 2004; Kaiser and Wright, 2006). Since the amount of a web page is hardly measurable due to its unlimited size unlike the page number in a magazine, I decide to use content variety for the measure. I confirmed that the number of sections from ResearchAd shows good variations according to user visits (see Choi et al. (2012)).
days. This average contract period is used as an instrument for the estimation. The justification of its exogeneity is discussed in the estimation section.

Number of unique visits for each website is shown in Table 1. One can check that two largest websites in the user market are Naver.com and Daum.net having more than 26 millions of user visits per month. Empas.com and Paran.com are the smallest firms with about one half of visits by Naver.com.

Six dummy variables are generated to capture the websites' observable characteristics. \(t_{elecom}\) represents if the website is owned by a telecommunication company. Nate.com and Paran.com are owned by SK telecom and KT, respectively, that are two biggest mobile companies in South Korea. \(t_{mail}\) is one if the website provides large capacity (more than one giga byte) e-mail service. Three of six websites provide \(t_{mail}\) service and two of them launched this service during the study period. \(t_{ecom}\) is one if the website provides e-commerce service. Three websites owned this service during the study period. \(t_{game}\) and \(t_{comm}\) is about offering gaming services and community services respectively. \(t_{comm}\) service is a grouping service that allows exclusive networking among members. \(t_{mhpg}\) is a mini-homepage service that helps users to have a personalized page and to network with pages of others (quite similar to Facebook.com today). This service was a big hit in Korea and had been popular until Facebook enters the Korean market.

The vector of exogenous characteristics, \(X_{jt}\) in advertising demand includes the number of sections, six dummies of website characteristics, the age of the website, the age squared, and constant.

Individual level website usage data

For the user side market data, I use visiting information for the six search engines during the study period. This data is gathered by Nielsen Korea. The Nielsen company puts
together a panel group for this survey. Each panel member has a specialized application running in her/his web browser so that the company can track when and how long she/he accesses a certain website. This data provides individual choice information for the website as well as demographics such as gender, age, education, income, jobs, and regions. The information on users is given for each choice of websites in each month. Table 2 presents what information I have at an individual level. Observations are per individual/website/month. They are 1,127,546 observations in total. I simulate the attention level of each user with this individual-level information. I also construct the user demand for a website with a random coefficient model with the individual choice data.

Page view (PV) is a count (number of hits) that increases every time the user clicks on pages belong to a certain website. Average PV is highest on Nate.com and lowest on Empas.com during the study period (see Table 2). To some extent, PV represents a performance in the user market, but it depends on the structure of the website. That is, if a website requires many clicks by construction, PV must be high for that website. Duration means how long a user stays on a website. One index of a user’s loyalty on a website can be obtained by dividing duration by PV. For example, a user clicks on pages in Naver.com 673 times while she/he stays a little more than 5 hours on that website during a month. Therefore, I can calculate the duration per click as 32.87 seconds for Naver.com. Empas.com has the highest number of 35.85 and Daum.net has the lowest of 27.84 for duration per PV (see Table 2). Daily frequency (DF) is a count of days in a month that a user accessed a website. For example, users visited Nate.com 12.18 days in a month on average in Table 2. DF can be a good measure for user visits because PV is somewhat misleading for the reason explained above.

Finally, the user behavior vector \( H_{jkt} \) has two indices for measuring how loyal a user is to a website: duration per PV and DF. The former represents how long a user stays
once she/he accessed a website and the latter measures how frequently she/he visits in a month. I made a vector using these two measures to describe user behavior. Interestingly, these two measures show different behavior depending on the websites: Empas.com and Paran.com have high average duration per PVs but low average DFs.

[Table 2 about here.]

Nielsen also provides demographic information of each panel member. I use four variables explaining the behavior in this study: gender, age, education level, and income. Other demographic information such as jobs and regions is used as excluded instruments. As shown in Table 2, there are almost no differences in three variables among the six websites. Average age and education level of users in the six websites show less than 1 year difference and for average income there is about a 50 US dollars difference between the lowest and the highest. This similarity among websites implies that these websites are general purpose websites. In other words, it is not likely to think that these websites target a certain group of consumers; they try to draw large fractions of people just as public television stations.

3.3 Estimation

Parameters in the ad demand function are estimated through a generalized methods of moment (GMM) procedure. I first present the formulation of the attention span and then describe how to derive moment conditions from the advertising demand model. For user demand, I apply maximum likelihood estimation to the mixed-logit model with individual choice data.
Attention level of each user

The determinants of the attention span are the user’s inherent characteristics and the actual exposure on the website. The demographic information (e.g. gender, age, and education level) and individual unobservables (e.g. inborn talents) constitute the users’ characteristics. Exposures can be measured by user behaviors (a duration of stay and a frequency of visits). I assume that the attention span \( m \) is positive and let \( m_{jkt} \) be the realization of the random variable \( m \) for user \( k \) on website \( j \) at time \( t \). Here, \( k \in \{1, \ldots, K\} \) is an index for each user. Therefore, \( m_{jkt} \) in the function \( g_{jt} \) represents the heterogeneity of individual attention in my model (see Eq. (2.3)). I assume this variable to be dependent on individual characteristics and behaviors. Specifically, \( m_{jkt} \) function satisfies the following:

\[
\ln m_{jkt} = D_k' \alpha^m + H_{jkt}' \beta^m + c^m + \zeta_{jkt}.
\] (3.1)

where \( D_k \) is a vector of user \( k \)'s characteristics and \( H_{jkt} \) is a vector of user \( k \)'s behavior on website \( j \) at time \( t \). \( c^m \) is a constant and \( \zeta_{jkt} \) is an error term. User behavior \( H_{jkt} \) is a \((2 \times 1)\) vector which represents user \( k \)'s time spent on the website and the number of access (frequency of visits) during period \( t \). The indicators in \( H_{jkt} \) are used to capture the variation across websites in the attention span of a given individual (demographic information remains the same across websites over time). For example, if a user stays longer and visits more frequently the website than others, this behavior can explain that more attention has been paid to the website. Likewise, these indices can represent the intensity of behavioral characteristics that a simple number of visits cannot explain. \( \zeta_{jkt} \) captures the impact of unobservable characteristics on the attention level. \( \alpha^m \) and \( \beta^m \) are parameters to be estimated.
There are two important issues in formulating this attention level equation. The first issue is that the attention spans are not observable. Therefore, I generate $m$ for each user. I simulate $m_{jkt}$ for all users given the distribution of $\zeta_{jkt}$ and Eq. (3.1). The second is the potential correlation between $H_{jkt}$ and the random term $\zeta_{jkt}$ in the $m$ function. More specifically, the person with higher $\zeta$ could possibly show different behavior from others on a website. The higher $\zeta_{jkt}$ could lead to longer duration or more frequent visits, or the other way around. I introduce additional equations to deal with this possible correlation as following:

$$H_{jkt} = \begin{pmatrix} D_k^h \alpha^d + \beta_1^d A_{jt} + X_{jt}^{h'} \beta_2^d + c^d + \epsilon^d_{jkt} \\ D_k^f \alpha^f + \beta_1^f A_{jt} + X_{jt}^{h'} \beta_2^f + c^f + \epsilon^f_{jkt} \end{pmatrix} \tag{3.2}$$

where $D_k^h$ is a vector of demographics of user $k$ and $X_{jt}^h$ is a vector of exogenous characteristics of website $j$ at time period $t$. $\epsilon^d_{jkt}$ and $\epsilon^f_{jkt}$ are the unobservable error terms which affects user $k$’s behavior. $\alpha^d$, $\beta^d$, $c^d$, $\alpha^f$, $\beta^f$, and $c^f$ are parameters to be estimated. Eq. (3.2) implies that an outcome of user behavior on website $j$ depends on individual characteristics, ad quantity, and website characteristics (including contents).

My estimation strategy consists in regressing $H_{jkt}$ separately on related variables and putting the fitted residuals as explanatory variables in Eq. (3.1). This is called control function method (Smith and Blundell, 1986). This ensures that the parameter estimates for the impact of $H_{jkt}$ on $m_{jkt}$ are consistent. Therefore, the attention span equation can be rewritten as following:

$$\ln m_{jkt} = D_k^h \alpha^m + H_{jkt}^{h'} \beta^m + c^m + \hat{E}_{jkt}^{h'} \sigma^h + \sigma^m c_{jkt}, \tag{3.3}$$
where $\hat{E}^h_{jkt} \sigma^h = \sigma^d\epsilon^d_{jkt} + \sigma^f\epsilon^f_{jkt}$ is a fitted value of residuals from the estimation Eq. (3.2). $\zeta_{jkt} = \hat{E}^h_{jkt} \sigma^h + \sigma^m\epsilon_{jkt}$ is, therefore, an error term of $m_{jkt}$ conditional on $\epsilon^d_{jkt}$ and $\epsilon^f_{jkt}$.

In this way, I take $m_{jkt}$ as (hypothetical) attention span of individual panel members. Then, I can determine whether each individual has a high enough attention level to process all messages received, which makes an individual acceptance ratio (or message processing ratio). $g_{jt}$ is computed by averaging these acceptance ratios.

Finally, the direct effects from the aggregate characteristics are assumed to be common across the individuals who have accessed the same website in the same period. This is why I didn’t include aggregate characteristics of websites such as $A_{jt}$ and $X_{jt}$ in the $m_{jkt}$ function. However, I can capture the effect indirectly from the user behaviors: users respond to the website characteristics first and then the resulting individual behaviors affect the attention span. There are also additional variables used only in the behavior equations for the identification such as the squared age of the website and the occupation and regional dummies of users.

**Advertising demand estimation**

The econometric specification of advertising demand is obtained by taking the logarithms of both sides of Eq. (2.2).

$$\ln P_{jt} = \alpha^p \ln A_{jt} + \beta^p \ln \phi_{jt} + X_{jt} \eta + v_{jt},$$

(3.4)

where $\alpha^p = \alpha - 1$ and I expect $-1 < \alpha^p < 0$ and $\beta^p > 0$ in the estimation. $v_{jt}$ is an error term. The observable characteristics ($X_{jt}$) and unobservables ($v_{jt}$) captures the effect of the logarithm of $\alpha \bar{\pi}_{jt}$ in Eq. (2.2).
The main issue of identification in Eq. (3.4) is to estimate separately the parameter $\beta^p$ and parameters of function $m_{jkt}$ in the expected number of ad noticed $\phi_{jt}$. It is possible due to the heterogenous distribution of the attention span of users. In other words, both types of users exist in the market: those who have large attention span enough to process all the ads received and the others who suffer from information congestion. The variation in the number of users who have $m_{jkt} \geq A_{jt}$ makes it possible to identify the parameter $\beta^p$. Likewise, the parameters in the $m_{jkt}$ function can be identified by the variation among characteristics of users who have $m_{jkt} < A_{jt}$ and by the nonlinear functional form.

[Table 3 about here.]

Endogenous variables in the inverse demand of advertising, Eq. (3.4), are $A_{jt}$ and $\phi_{jt}$. As the model says, the ad quantity and the expected users noticing the ad explain the ad price and, inversely, are explained by unobservable quality factors that change the ad price (e.g. server network capacity, the provision of popular contents, or so), implying the possible endogeneity. I select proper exclusive variables for GMM instruments to deal with this endogeneity.

Excluded instruments can be divided into three groups: average behavior of advertisers, average characteristics of users, and overall market condition. As the advertiser’s behavior I include the average contract periods. To use this variable as an instrument, I need to make an assumption that the contract period is exogenous to the ad demand. This is counterintuitive because normally the advertising quantity has an inverse relationship with the length of the contract period. However, the average contract periods on websites are around a week (see Table 1) and the variation is closely correlated with the firm’s characteristic that is exogenous. From the contingency table of the contract period and the industrial category, I can confirm that two variables are not independent.
and have a high correlation (Pearson’s Chi-Square test statistic is larger than 3,200). In addition, the contract period is not significantly correlated with the ad quantity (see Table 3).

The average contract periods of rival platforms is used to instrument the number of users. It is because the rivals’ decision of the ad quantity would affect my user demand but by construction it is independent from my decision of ads. Therefore, the contract periods of rivals affects the user demand but it is exogenous to the ad demand. I choose some average profiles of users as the excluded instruments: gender, student, age, education, income, marriage, and region. These average characteristics are assumed to be exogenous to the decision of the advertising supply. This is the same logic as in the exogeneity of the content amount. The last excluded instrument is the total internet users in the market that shows the overall market condition.

Table 3 shows the result of the pseudo first-stage estimation with regard to the ad quantity and the user visit variables in the advertising demand function. For the ad quantity variable, most included variables such as website characteristics are significant and 57% of variations are explained ($R^2 = 0.5695$). The number of user visits are significantly explained by most excluded variables and user characteristics and the corresponding $R^2$ is 0.9254. Notice that the average contract periods doesn’t have significant correlation with the aggregate ad level but with the user visits. This is also the same for the average of rivals’ contract periods, supporting the property as instruments explained above.

The panel property of my data is also applied to the identification. Exogenous characteristics of the websites have variations over time, though not much. Continuous variables such as number of sections, age of websites, and total internet users change over time. In addition, dummy variables such as lmail and mhpq are not constant because these services are newly introduced in the middle of the study period. It is comparable to the case of Nevo (2001) where the exogenous variables are constant for the whole observation
period.

I derive a residual for the moment condition of ad demand from Eq. (3.4) as $\nu_{jt} = \ln P_{jt} - \alpha^p \ln A_{jt} - \beta^p \ln \phi_{jt} - X_{jt} \eta$. To compute the residual $\nu_{jt}$, I need the value of the expected number of users who notice the ad, $\phi_{jt}$, and thus, the individual’s attention level. Then, a set of parameters $\{\alpha^p, \beta^p, \eta, \alpha^m, \beta^m, c^m, \sigma^m, \sigma^d, \sigma^f\}$ is chosen in order that they minimize the GMM objective function, $\Lambda'_A Z_A W^{-1}_A Z'_A \Lambda_A$, where $\Lambda_A = \nu$ is an error term, $Z_A$ is a vector of instruments. $W_A$ is a weight matrix that is a consistent estimate of $E[Z'_A \Lambda_A \Lambda'_A Z_A]$.

In order to improve the efficiency and the speed of the estimation, I applied two techniques in the estimation process: importance sampling and antithetic acceleration. To construct the importance sampler, I draw more samples from the users who are highly probable to face information congestion. The probability of being congested is computed with the result from the first round estimation that is done without any prior information. The antithetic acceleration is a simple method which can increase the efficiency of the simulators (Stern, 1997). More detailed explanations can be found in the Appendix.

**User demand estimation**

Thanks to the individual choice information, I can utilize a discrete choice model, specifically mixed-logit, for the user demand and I use maximum likelihood estimation (MLE) with simulation methods (Train, 2003) for estimating the parameters. The simulated choice probability of Eq. (2.5) is constructed as follows:

$$P_{jkt}^R = \frac{1}{R} \sum_{r=1}^R \frac{e^{V_{jkt}(\theta^r_k)}}{\sum_{l=1}^J e^{V_{jkt}(\theta^r_k)}},$$

where $R$ is the number of draws and $\theta^r = \{\rho^r_k, \lambda^r_k\}$ is $r$th draw of random coefficients.
from the cumulative probability distribution \( G(\cdot) \). \( V_{jkt}(\theta^e_r) = \rho_k^e A_{jt} + X_{jt} \lambda_k^e + c^e_j \) is a mean utility of \( k \) choosing \( j \) where \( \rho_k^e = \rho + D_{kt} \omega + \sigma^p \mu_{jkt}^p \) and \( \lambda_k^e = \lambda + \sigma^\lambda \mu_{jkt}^\lambda \). I assume that \( \mu_{jkt}^p \) and \( \mu_{jkt}^\lambda \) are identically and independently distributed and following standard normal distribution. \( \{\rho, \sigma^p, \lambda, \sigma^\lambda, c^e_j\} \) are parameters to be estimated. Here, I include the choice-specific constant only for the outside option, that is, \( c^e_{j} = 0 \), \( j \in \{1, ..., J\} \) and \( c^e_0 \neq 0 \). The reason is that by construction all the choice-specific constants cannot be identified in logit models and so I let \( c^e_0 \) represent the fixed-effect on the utility of the outside option relative to the options in \( \{1, ..., J\} \) that I assumed to have the same fixed-effects. Besides, I assume that the exogenous characteristics of each website could capture the unobserved website-specific effects that might change over time.

One issue here is that the sampling method is not exogenous, but it is choice-based. I sample the same number of individuals for each choice. The simple estimation without taking into account this issue can lead to a sampling bias. I have the aggregate market share information of the websites in each period, and so I apply this information as a weight when estimating the individual choice probabilities (Manski and Lerman, 1977).

The simulated log-likelihood (SLL) function is defined as \( \sum_{t=1}^{T} \sum_{k=1}^{K} w_{jt} \ln P_{jkt}^R \), where \( w_{jt} \) is the aggregate market share of website \( j \) in period \( t \). The idea is that \( w_{jt} \) gives more weight to sample log-likelihoods from the website with higher market share and vice versa. Then, the maximum simulated likelihood (MSL) estimator is a set of parameters that maximizes the SLL. The property of MSL estimator is known to be consistent, asymptotically normal, and efficient if \( \sqrt{N/N} \to 0 \) with sample size \( N \) (Train, 2003).
4 Result

4.1 Parameter estimates of the model

The estimation results of the advertising demand are shown in Table 4. I consider two different specifications of the estimation. Specification (I) is the result of the demand function without information congestion. Specification (II) is the result from the model with information congestion. Parameters of interest, $\alpha^p$ and $\beta^p$, are shown to be consistent with the theory. The signs of $\alpha^p$ in all two specifications are shown to be negative as I expected. The estimate of $\alpha^p$ in specification (I) has larger size than the estimate in specification (II). This may imply that negative slope of demand could be overestimated without considering information congestion effect. The effect of expected ad noticed, $\phi_{jt}$, is significantly positive. This shows that increasing expected views on a website raises the price, showing the positive network externality from the user side. The condition for Proposition 1 ($1 - \hat{\alpha}^p < \hat{\beta}^p$) is also confirmed by the estimation result.

[Table 4 about here.]

The number of sections that approximates the content amount has positive effect on the ad demand. This tells me that advertisers value the number of sections on the websites. The possible reason for this is that more sections mean more provision of the ad spaces. In addition, having more sections could be a signal that the website has more capability to invest in contents and thus, to draw users. Among dummy variables, parameters for $lmail$, $ecom$ and $comm$ have negative signs and the others have positive ones. The negative effect of large e-mail service might imply its low contribution to the display advertising or the unsuccessful linkage of e-mail services with the advertising. In the same way, e-commerce and community services might not have been effective for the advertising delivery. Gaming services on websites are usually online services that
can make connections among multiple players and *mhpg* are a type of social networking services like Facebook. Therefore, *game* and *mhpg* have an important feature in common: they keep users connected. The positive signs of those two parameters can imply that the connection among users might have produced good outcomes for advertisers. The estimate of *telcom* dummy is significant only in specification (II) and it is negative. The reason might be that the websites belong to the large telecom companies are not the first movers but followers so they have had a certain disadvantage in the market.

The result from the attention level equation is presented in the last column of Table 4. All parameter estimates are shown to be significant. A dummy for male users has negative sign implying that they have shorter attention spans than females do. Parameter for the age and the age squared show that the attention span increases with the age. The positive sign of education level says that people with higher education level could pay more attention to the ad messages. The effects of duration and frequency on the attention level is significant and negative. This seems surprising and the reason is the inclusion of the control functions from the user-behavior equations. The negative sign of user behaviors on the attention implies that the person with low attention level would have stayed longer and visited more frequent.

[Table 5 about here.]

In order to produce the error terms $\epsilon^d$ and $\epsilon^f$ as the control functions, I run the user behavior equation of Eq. (3.2) first. The estimation results of Eq. (3.2) are shown in Table 5. The parameters for the ad quantity are negatively significant and the parameters for the content amount are positively significant in *duration* and *frequency*. This implies that the content can keep users visit and stay but ads work in the opposite way as I expected.
The result from the user demand estimation is shown in Table 6. Random coefficients are assigned to the ad quantity and the number of sections. The resulting parameter distribution for the ad quantity is shown to have the estimated mean of -.0038 and the estimated standard deviation of .0009 such that 92.2% of users have negative effect, i.e. nuisance, for the increase of advertising. Among interactions of demographics with the aggregate ad level, user’s age and the education level are shown to be significantly positive so that younger and less educated users would be less attracted by the advertising. The interaction of the advertising with male users shows positive correlation but not significant. The mean effect of the number of sections is positive and significant. One can see that 98.9% of users have positive coefficient on the number of sections. Estimates of three aggregate dummies are significant: telcom, ecom, and mphg, meaning that the site’s affiliation under telecom companies, large e-mail service, and mini-hompage service give positive impact on the choice probability. Site’s age also shows positive and significant relationship with the user demand, meaning that websites with more experiences could be valued by the users. The constant for the outside option is estimated negative and significant showing that it has relatively negative fixed-effect compared to the websites in the choice-set.

4.2 Message processing rates

Main estimate of interest would be the probability $g_{jt} = \int \min\left(1, \frac{m}{A_{jt}}\right) dF(m)$ in each website. This value can be computed by averaging message processing rates of each user. I present the values of $g_{jt}$ in Table 7. Estimated values are different according to websites: the probabilities range from .36 to nearly .67, where the median for each website can be as low as .44 and as large as .54. The variation of these rates over
websites are interesting. Seemingly, websites with low market share such as Empas.com and Paran.com has higher rates than those with high market share. The fact that advertisers suffer information congestion more in the popular websites might imply the possibility of exploitation of higher market power. This can be also shown with the elasticity estimates.

4.3 Elasticities with respect to advertising change

I present elasticities of prices with respect to ads in the Table 8. These values are computed numerically as the amount of price changes by the 1% change of ad level. First column is about elasticities under the assumption that information congestion model, i.e. specification (II) is the true one. The second column is for the assumption that specification (I) is the true model.

These elasticities are also interpreted as negative values of Lerner index, that is, markups of platforms in the advertising market. Overall, the absolute sizes are bigger in the case with information congestion. This implies that the estimation of markups could be biased without considering information congestion. Specifically, ignoring information congestion, I possibly underestimate the size of markups. I also compute marginal costs using these elasticities. Estimated marginal costs are shown in Table 9, where the average marginal cost is about 4,800,090 won in specification (II) and about 8,615,640 Won in specification (I). This shows that if one doesn’t consider the effect of information congestion, marginal costs can be overestimated.
In the third column in Table 8, I show the change in the user shares by 1% rising in the ad level. An increase in ad level makes some users to opt out due to the ad nuisance. On average, .39% of users are shown to switch to others when 1% of ad level rises.

The elasticity of $g_{jt}$ with respect to ads are shown in the fourth column in Table 8. One can see that the websites with high market share have high elasticity of message processing rate. More importantly, the markup is explained more by the scarcity of attention than by the network effect from user’s side. That is, the attention-advertising elasticity, $\varepsilon_A^g$, is higher than the users-advertising elasticity $\varepsilon_U^A$ on average ($\bar{\varepsilon}_A^U = -0.0039$ and $\bar{\varepsilon}_A^g = -0.0545$).

[Table 9 about here.]

4.4 Counterfactual: Effect of introducing an option to avoid ads

This subsection explores a simple counterfactual experiment on estimating the effect of introducing the ad avoiding option in a website. I assume that there is a subscription price for users who want to remove ads on the website.\textsuperscript{18} Previous studies on the effect of ad-avoidance predict that the ad-level increases with the introduction of the avoiding options. Tag (2009) explains that ad-level rises because of the second-degree price discrimination to differentiate the high-quality goods (without ads) from damaged goods (with ads). Another explanation is that the advertiser’s lower valuation of ad viewers with inelastic viewer demand causes the ad-level increase (Anderson and Gans, 2011; Wilbur, 2008).

In this experiment, I find the effect of the ad-avoidance on the information congestion as well as on the ad-levels and prices using the structural model in this paper. Given that

\textsuperscript{18}This should require users to sign-in on the website as an additional burden but I assume that this effect has already incorporated into the given price.
the total level of attention has decreased due to the ad-avoidance, the change in the ad-level would affect both the user demand and the information congestion but in different extent. I describe the experiment setting and the estimation result of this regime change.

I first assume that the share of ad-avoiders is common for all platforms and that it is given exogenously. Let $\tau$ be the share of ad-avoiders. Platforms set their subscription prices to meet this share. Let $P^S_j$ be a subscription price that allows users to avoid advertising. The rate of $\tau$ users choose to subscribe to the platform $j$ and don’t watch the advertising. The two different utility functions for the ad-avoiders and the ad-viewers are:

$$u_{jk} = \begin{cases} 
\delta P^S_j + X_j \eta + \epsilon_{jk}, & \text{if user } k \text{ chooses to pay the fee.} \\
\rho_k A_j + X_j \eta + \epsilon_{jk}, & \text{otherwise.} 
\end{cases}$$

where $\delta$ is a price coefficient and the subscription price is decided as the share of users who has lower $\rho_k$ (larger in absolute value) than $\tilde{\rho}_k$ would be $\tau$ where $\delta P^S_j = \tilde{\rho}_k A_j$.

Then, the number of users $U_j = \tau U^S_j + (1 - \tau) U^A_j$ where $U^S_j$ is the number of ad-avoiders and $U^A_j$ is the number of ad-viewers. Therefore, platform $j$’s problem becomes $\max_{A_j} \Pi_j = P_j (A_j, U^A_j) A_j + P^S_j U^S_j - c_j A_j$, where $P_j = (\alpha \hat{\pi}_j A_j)^{\beta} \left[ g(A_j) U^A_j \right]$. The first-order condition is as following:

$$P_j - c_j + \varepsilon_A^P P_j + \varepsilon_P^P \varepsilon_A^g \frac{P_j}{A_j} + \varepsilon_P^P \varepsilon_U^A \frac{P_j}{U_j} = 0,$$

(4.1)

where the elasticities $\varepsilon_A^g$ and $\varepsilon_U^A$ are re-estimated using $U^A_j$. I compute the equilibrium ad-levels, prices, and the average attention spans that satisfy the condition in Eq. (4.1).

[Figure 2 about here.]
The results are shown in Figure 2. When the share of ad-avoiders is smaller than .4 the ad-levels are higher than the initial ad-level with no ad-avoiders. After then, it decreases in the share of ad-avoiders. This result is opposite to the ones from the previous studies. Then reason might be that the value of ad-viewers are increasing when the users with high ad nuisance switch to avoid ads. To some extent, the graph in the bottom of Figure 2 can explain the reason. The average attention level doesn’t change much with the decrease in the total attention level. This reflects the high elasticity of attention according to the ad avoiders so that the price of attention gets higher. Still, the overall revenue of a platform is decreasing and the ad-avoiders can enjoy the surplus with the fixed subscription price so the introduction of this option is not much profitable for platforms under the restrictive assumption in this experiment.

5 Conclusion

Website users are heterogenous in accepting the advertising messages. I consider this heterogeneity as a model of the information congestion in the advertising demand. I follow Rysman (2004) in building traditional two-sided market structure, and extend the model with the specification of the information congestion. I show that there is a possibility of bias in the demand estimation without considering the advertising congestion. The attention equation formulates the relationships between the website users’ characteristics and the attention level. I estimate the average attention span and show its variation by websites. The individual choice data allows me to employ the mixed-logit model for the user demand estimation. The result shows that 92.17% of users have the nuisance effect on the advertising level. The counterfactual experiment shows how the share of ad-avoiders affects the equilibrium ad-level and the average attention level on the websites.
As a limitation of the research, the assumption that a user makes a discrete choice for a website (i.e. assuming a single-homing user) is crucial in my model. Although the single-homing assumption is usual in the conventional media economics, users can choose multiple websites in reality. Therefore, one interesting extension will be exploring the multi-homing behavior by consumers and find out the resulting effects in both sides of the market.
Appendix: Information congestion in various media types

Based on the applications in the previous studies, I compare the attention-giving behaviors of consumers depending on the various types of advertising and media in Table 10. The advertising in direct mails and telemarketing sends messages directly from the advertisers to the consumers while others send messages through media platforms. Two main characteristics can be considered in describing these advertising types: targeting (if the consumers are targeted) and soliciting (if the consumers solicit the messages). Direct mails, telemarketing, and magazine ads employ targeting technology. Direct mails and telemarketing use the personal information to target the favored consumer groups (see Anderson and de Palma (2009) and Van Zandt (2004)) and magazines focus on a certain topic or interest to draw the attention of the related group of people (see Kaiser and Wright (2006)). Other platforms in Table 10 are considered as the generalized platforms rather than the targeted ones.

The consumer’s attitude towards ads depends on whether the ad messages are solicited or not. It is positive for the solicited advertising but negative for the unsolicited advertising. Ad messages in Yellow Pages can be seen as the solicited advertising. In other words, the readers of Yellow Pages are motivated to find the information in the book. Rysman (2004) shows that the effect of advertising on the reader’s demand is positive in Yellow Pages markets. In contrast, the unsolicited messages can cause annoyance for the consumers due to the unexpected signals (in stage 1). This requires the consumers to decide whether they accept the messages or not (in stage 2). Consumers are usually passive and show negative reactions (nuisance effect) to these unsolicited ad messages. The parameter estimates of the advertising in the consumer demand are shown to be negative in television networks (Wilbur, 2008) and websites (Choi et al., 2012). The information congestion arises if the consumers discard the messages in this stage. The viewers in Yellow Pages and magazines don’t feel nuisance about ad messages and so...
they don’t need to go through the stage 2 in Table 10. Finally, what advertisers want here is the contact from the potential consumers as shown in stage 4. Advertisers can expect the increase in their profits when the more consumers attain this stage.

The important implication in Table 10 is that the information congestion can happen in different stages: when the consumers decide not to respond to signals (in stage 2) or when the consumers give up to contact advertisers (in stage 4). The advertising congestion occurring in the later stage is more likely to be called as “business stealing effect” which takes place among the firms in the same product category (see Rysman (2004) or discussions in Anderson and de Palma (2012, 2013)). My model focuses on the congestion problem in the former stage while Rysman (2004), for example, deals with the congestion in the latter stage. The online display advertising from diversified industries reach to consumers so the business stealing effect like in Yellow Pages is less likely to happen.

[Table 10 about here.]

B Appendix: The message processing probability function

B.1 Derivation of \( g_{jt} \)

I consider “number of message losses out of \( A_{jt} \) given users in \( j \) at period \( t \)”, \( \text{Loss}(A_{jt}) \), as following:

\[
\text{Loss}(A_{jt}) = F(A_{jt}) \int_0^{A_{jt}} (A_{jt} - m) \frac{f(m)}{F(A_{jt})} \, dm + (1 - F(A_{jt})) \times 0,
\]

where \( m \) is a random variable whose cumulative distribution function is \( F(\cdot) \) and the
density is \( f(\cdot) \). The former term in the RHS is the message losses when the attention span \( (m) \) is smaller than the total messages \( (A_{jt}) \) and the latter term is zero when the attention span is bigger than the total messages. Message loss rate would be, \( \frac{\text{Loss}(A_{jt})}{A_{jt}} = \int_0^{A_{jt}} \left( 1 - \frac{m}{A_{jt}} \right) f(m) \, dm \), so the message processing rate, \( g_{jt} \) is:

\[
\begin{align*}
g(A_{jt}) &= 1 - \frac{\text{Loss}(A_{jt})}{A_{jt}} = 1 - \int_0^{A_{jt}} f(m) \, dm + \int_0^{A_{jt}} \frac{m}{A_{jt}} f(m) \, dm \\
&= 1 - F(A_{jt}) + \int_0^{A_{jt}} \frac{m}{A_{jt}} f(m) \, dm .
\end{align*}
\]

Here, the formulation says that each user suffers information congestion when her attention span is less than the total ad messages in website \( j \) at time \( t \) (i.e. \( m < A_{jt} \)).

An important aspect of user behavior on examining display ads is that the average expectation on the remaining messages is the same as the \textit{ex ante} expectation. The total numbers of messages and processes are, therefore, important in formulating the message process rate. This is as discussed in Van Zandt (2004) and Anderson and de Palma (2009) where each individual accepts \( \min \left( 1, \frac{m}{A_{jt}} \right) \) of total messages.

\[\text{B.2 Proof of the differentiability of } g_{jt}\]

I suppress the subscripts \( j \) and \( t \) for the exposition purpose. The message processing rate \( g(A) \) can be decomposed as following:

40
\[ g(A) = \int_0^\infty \min \left( 1, \frac{m}{A} \right) f(m) \, dm \]
\[ = \int_0^A \frac{m}{A} f(m) \, dm + \int_A^\infty f(m) \, dm \]
\[ = \int_0^A \frac{m}{A} f(m) \, dm + 1 - F(A), \]

where \( m \) is a random variable whose cumulative distribution function is \( F(\cdot) \) and the
density is \( f(\cdot) \). I assume that \( F(\cdot) \) and \( f(\cdot) \) are twice differentiable functions. Then, the
derivative of \( g(A) \) with respect to \( A \) can be shown as following:

\[ \frac{dg}{dA} = \frac{A}{A} f(A) - \int_0^A \frac{m}{A^2} f(m) \, dm - f(A) \]
\[ = - \int_0^A \frac{m}{A^2} f(m) \, dm. \]
C Appendix: Simulation techniques and estimation procedure

I simulate individual attention spans to build a moment condition of advertising demand. Two simulation techniques are applied in the process: importance sampling and anti-thetic acceleration. I give brief explanations on their implementation in this Appendix.

C.1 Importance sampling

I employ this technique in order to smooth \( g(A_{jt}, m_{jkt}) = \sum_k^n \min \{ 1, \frac{m_{jkt}}{A_{jt}} \} \) function. The \( \min \{ \cdot \} \) function here restricts the congestion ratio to 1 if the attention span \( m_{jkt} \) exceeds the ad level \( A_{j} \), and this makes it difficult to estimate parameters in \( m_{jkt} \). Therefore, it would be better to have more samples who have lower attention span than the ad level (just for the sake of numerical estimation).

The idea of this technique is that I draw more from “congested” part of samples than the other. To do that, I need a prior information about samples. I run the first stage estimation with regular draws. Using estimates of \( \theta_1 \) from the first stage, sample again with the following probability:

\[
Pr(A_{jt} > m_{jkt}|\theta_1) = \Upsilon \left( \frac{A_{jt} - D_k \alpha^m - \beta^m h_{jkt}}{\sigma^m} | \theta_1 \right),
\]  

where \( \Upsilon(\cdot) \) is a cdf of log-normal distribution. If this probability is high with a sample, it is more likely to be drawn. Like this way, I construct the sample draws again. Then, I re-do the estimation process with this sample draws. In this stage, I should give weights corresponding to the inverse of this probability to each sample. It is simply because I draw more from the “congested” part of samples, so I give less weights to them.
C.2 Antithetic acceleration

Antithetic acceleration method is used to speed up the simulation process and to acquire more stable results by reducing variances. This applies to the $g_{jt}$ function where I average the samples drawn. $g_{jt}$ function is originally given as follows:

$$E_k \left[ \min \left\{ 1, \frac{m_{jkt}(\zeta_{jkt})}{A_{jt}} \right\} \right] = \frac{1}{ns} \sum_k^{ns} \min \left\{ 1, \frac{m_{jkt}(\zeta_{jkt})}{A_{jt}} \right\},$$

where $\zeta_{jkt}$ is an i.i.d. draw from log-normal distribution. The way how antithetic acceleration is applied is that I sum the additional random draws from the opposite part of the distribution. If the random draws follow standard normal distribution $N(0, 1)$, then I add negative value of draws from the same distribution. This would produce the same average value, but reducing extremities. When one draw is unusually large, then the other one is unusually small so that two extremities will be averaged out (Stern, 1997). Therefore, the following formulation will produce the same $g_{jt}$ as the former, but will be more efficient:

$$\frac{1}{2 \times ns} \sum_k^{ns} \left[ \min \left\{ 1, \frac{m_{jkt}(\zeta_{jkt})}{A_{jt}} \right\} + \min \left\{ 1, \frac{m_{jkt}(\zeta_{jkt}^{-1})}{A_{jt}} \right\} \right], \quad (C.2)$$

where I sum additional draws from $\zeta^{-1}$ since $\zeta$ follows log-normal distribution.

C.3 Estimation procedure

Generalized method of moments (GMM) is a good way of estimating parameters in non-linear models. GMM has good large sample properties. Also, it is easy to implement and to achieve convergence. For an efficient GMM estimation, I perform the estimation
process with an assumption of homoscedastic error terms in the first stage (just applying $Z'Z$ as a weight), and then, I re-do the process with the weight computed by $\hat{\Lambda}$ using estimated parameters in the first stage. Besides, I compute probabilities for applying importance sampler with the result from the first stage. I present the brief description of the estimation process below:

1. Initial random draws are prepared for the estimation: $\zeta_{jt}$ and draws from user samples. I do not change these draws throughout the whole estimation process.

2. For chosen parameter values and sample draws, I compute $m_{jkt}$.

3. I compute $\phi_{jt}$. Antithetic acceleration is applied when calculating $g(\cdot)$ function,

$$\frac{1}{2\times ns} \sum_k^{ns} \left[ \min \left\{ 1, \frac{m_{jkt}(\zeta_{jkt})}{A_{jt}} \right\} + \min \left\{ 1, \frac{m_{jkt}(\zeta_{jkt}^{-1})}{A_{jt}} \right\} \right].$$

4. For given values of $\alpha^p$ and $\beta^p$, derive residual of inverse demand function, $\upsilon_{jt}$.

5. I evaluate GMM objective function with weighting matrix, $W$. (when initial stage, use $E[Z'Z]$, assuming homoscedasticity. It is consistent but not efficient.)

6. When the initial process converges, I draw samples again with the probability of $Pr(A_{jt} > m_{jkt}|\theta_1)$ which is calculated with the initial estimates.


8. Iterate from step 2 to 5 until finding parameters that minimize GMM objective given a tolerance level. The inverse of the sampling probability should be applied when computing $m_{jkt}$ (importance sampler).
References


Figure 1: Revenue Growths of Banner Ads in Six Search Engines.
Figure 2: Change of the ad-level and the average attention level according to the share of ad-avoiders.
Table 1: Description of Aggregate-Level Variables (Monthly Averages).

<table>
<thead>
<tr>
<th></th>
<th>Naver</th>
<th>Daum</th>
<th>Nate</th>
<th>Yahoo</th>
<th>Empas</th>
<th>Paran</th>
<th>Pooled(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Price (in dollars)</td>
<td>16,022.39</td>
<td>15,728.24</td>
<td>25,081.58</td>
<td>30,025.09</td>
<td>9,323.36</td>
<td>15,308.55</td>
<td>18,581.54</td>
</tr>
<tr>
<td>Ad Quantity</td>
<td>633.91</td>
<td>572.39</td>
<td>334.94</td>
<td>205.21</td>
<td>95.91</td>
<td>114.39</td>
<td>326.13</td>
</tr>
<tr>
<td>N. of Advertisers</td>
<td>148.242</td>
<td>101.758</td>
<td>60.485</td>
<td>41.485</td>
<td>42.788</td>
<td>27.939</td>
<td>70.449</td>
</tr>
<tr>
<td>Unique Visits</td>
<td>26,840,818</td>
<td>26,180,139</td>
<td>23,630,778</td>
<td>20,316,316</td>
<td>13,100,898</td>
<td>14,590,839</td>
<td>20,776,631</td>
</tr>
<tr>
<td>ARPU(^b) (in dollars)</td>
<td>0.235</td>
<td>0.211</td>
<td>0.130</td>
<td>0.113</td>
<td>0.058</td>
<td>0.068</td>
<td>0.136</td>
</tr>
<tr>
<td>Obs.</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>198</td>
</tr>
</tbody>
</table>

\(^a\) Results from pooling all observations.

\(^b\) Average Revenue Per User.
Table 2: Website usage and demographics (individual level, monthly averages).

<table>
<thead>
<tr>
<th></th>
<th>Naver</th>
<th>Daum</th>
<th>Nate</th>
<th>Yahoo</th>
<th>Empas</th>
<th>Paran</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Views (PV)</td>
<td>673.03</td>
<td>799.90</td>
<td>954.66</td>
<td>219.93</td>
<td>126.61</td>
<td>157.03</td>
</tr>
<tr>
<td>Duration (in seconds)</td>
<td>18,736.92</td>
<td>19,612.01</td>
<td>20,801.21</td>
<td>5,743.17</td>
<td>3,643.36</td>
<td>2,945.23</td>
</tr>
<tr>
<td>Duration / PV</td>
<td>32.87</td>
<td>27.84</td>
<td>30.66</td>
<td>31.96</td>
<td>35.85</td>
<td>34.94</td>
</tr>
<tr>
<td>Daily Frequency (DF, in days)</td>
<td>13.28</td>
<td>12.71</td>
<td>12.18</td>
<td>6.74</td>
<td>5.43</td>
<td>4.92</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>33.36</td>
<td>33.50</td>
<td>33.18</td>
<td>33.40</td>
<td>33.68</td>
<td>33.73</td>
</tr>
<tr>
<td>Education (in years)</td>
<td>11.68</td>
<td>11.72</td>
<td>11.80</td>
<td>11.64</td>
<td>11.91</td>
<td>11.84</td>
</tr>
<tr>
<td>Income (in 10 dollars)</td>
<td>357.78</td>
<td>357.15</td>
<td>355.72</td>
<td>360.99</td>
<td>362.07</td>
<td>360.18</td>
</tr>
<tr>
<td>Obs.</td>
<td>250,683</td>
<td>244,108</td>
<td>223,696</td>
<td>181,274</td>
<td>122,227</td>
<td>135,258</td>
</tr>
</tbody>
</table>

1,157,246 obs. in total.
Table 3: Pseudo first-stage estimations of the ad demand equation.

<table>
<thead>
<tr>
<th>Advertising Demand</th>
<th>$\log(AdQuantity)$</th>
<th>$\log(UserVisits)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. rivals’ contract days</td>
<td>-0.2562</td>
<td>(0.2031)</td>
</tr>
<tr>
<td>Avg. contract days</td>
<td>0.0039</td>
<td>(0.1097)</td>
</tr>
<tr>
<td>Total internet users</td>
<td>-0.00006</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Male</td>
<td>1.3032</td>
<td>(7.5742)</td>
</tr>
<tr>
<td>Student</td>
<td>-6.3078</td>
<td>(15.8981)</td>
</tr>
<tr>
<td>User’s Age</td>
<td>1.0838</td>
<td>(0.5606)</td>
</tr>
<tr>
<td>Education</td>
<td>-2.0819</td>
<td>(1.3776)</td>
</tr>
<tr>
<td>Income</td>
<td>0.0270</td>
<td>(0.0401)</td>
</tr>
<tr>
<td>Married</td>
<td>-21.457</td>
<td>(17.6236)</td>
</tr>
<tr>
<td>Seoul</td>
<td>-4.9476</td>
<td>(6.1341)</td>
</tr>
<tr>
<td>$\log(Sections)$</td>
<td>0.5693</td>
<td>(0.3214)</td>
</tr>
<tr>
<td>Website Age</td>
<td>0.7015</td>
<td>(0.2266)</td>
</tr>
<tr>
<td>Website Age$^2$</td>
<td>-0.0374</td>
<td>(0.0146)</td>
</tr>
<tr>
<td>$D_{mail}$</td>
<td>-0.2498</td>
<td>(0.2443)</td>
</tr>
<tr>
<td>$D_{ecom}$</td>
<td>2.1981</td>
<td>(0.0697)</td>
</tr>
<tr>
<td>$D_{game}$</td>
<td>-1.1386</td>
<td>(0.5898)</td>
</tr>
<tr>
<td>$D_{comm}$</td>
<td>2.4403</td>
<td>(0.7253)</td>
</tr>
<tr>
<td>$D_{mbpg}$</td>
<td>-1.5142</td>
<td>(0.5777)</td>
</tr>
<tr>
<td>$D_{telcom}$</td>
<td>2.6006</td>
<td>(1.0010)</td>
</tr>
<tr>
<td>Const.</td>
<td>-6.0229</td>
<td>(29.0370)</td>
</tr>
</tbody>
</table>

Adj. $R^2$ 0.5695 0.9254
Obs. 198

Note: This is OLS estimation that is separately done from GMM estimation. There is no actual first-stage estimation in GMM. Excluded instruments are shown in bold.
Table 4: GMM estimates of the advertising demand function.

<table>
<thead>
<tr>
<th></th>
<th>(I) Traditional Model</th>
<th>(II) Congestion model</th>
<th>(II) Congestion model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ad Quantity ($\hat{\alpha}_p$)</td>
<td>Expected User Demand ($\hat{\beta}_p$)</td>
<td>Num. of Sections</td>
</tr>
<tr>
<td>Ad Quantity ($\hat{\alpha}_p$)</td>
<td>-0.7778 (0.1237)</td>
<td>0.6631 (0.0405)</td>
<td>0.3618 (0.1902)</td>
</tr>
<tr>
<td>Expected User Demand ($\hat{\beta}_p$)</td>
<td>-0.6112 (0.0008)</td>
<td>0.6553 (0.0059)</td>
<td>0.4846 (0.0022)</td>
</tr>
<tr>
<td>Num. of Sections</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>Age of Website</td>
<td>Male (0.1682)</td>
<td>Age (0.1403)</td>
<td>$Age^2$ (0.1923)</td>
</tr>
<tr>
<td>$D_{tmail}$</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>$D_{ecom}$</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>$D_{game}$</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>$D_{comm}$</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>$D_{mhpg}$</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>$D_{telcom}$</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
<tr>
<td>Const.</td>
<td>-3.3172 (0.1682)</td>
<td>-3.3473 (0.1403)</td>
<td>1.4376 (0.1923)</td>
</tr>
</tbody>
</table>

Note: Standard errors are reported in parentheses.
Table 5: The additional estimations of user behavior equations.

<table>
<thead>
<tr>
<th></th>
<th>Log (Duration per PV)</th>
<th></th>
<th>Log (Frequency)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log(AdQuantity)</td>
<td>-0.0183</td>
<td>(0.0037)</td>
<td>-0.0207</td>
<td>(0.0063)</td>
</tr>
<tr>
<td>log(Sections)</td>
<td>0.0867</td>
<td>(0.0149)</td>
<td>0.4217</td>
<td>(0.0258)</td>
</tr>
<tr>
<td>Website.Age</td>
<td>0.0644</td>
<td>(0.0080)</td>
<td>0.3009</td>
<td>(0.0137)</td>
</tr>
<tr>
<td>Website.Age^2</td>
<td>-0.0045</td>
<td>(0.0006)</td>
<td>-0.0245</td>
<td>(0.0011)</td>
</tr>
<tr>
<td>D_email</td>
<td>-0.0020</td>
<td>(0.0096)</td>
<td>-0.2932</td>
<td>(0.0167)</td>
</tr>
<tr>
<td>D_email</td>
<td>-0.0344</td>
<td>(0.0308)</td>
<td>0.8581</td>
<td>(0.0532)</td>
</tr>
<tr>
<td>D_game</td>
<td>0.0967</td>
<td>(0.0289)</td>
<td>0.0355</td>
<td>(0.0499)</td>
</tr>
<tr>
<td>D_email</td>
<td>0.0777</td>
<td>(0.0334)</td>
<td>0.3961</td>
<td>(0.0577)</td>
</tr>
<tr>
<td>D_email</td>
<td>-0.0463</td>
<td>(0.0286)</td>
<td>0.0404</td>
<td>(0.0494)</td>
</tr>
<tr>
<td>D_telecom</td>
<td>-0.0130</td>
<td>(0.0391)</td>
<td>0.6573</td>
<td>(0.0675)</td>
</tr>
<tr>
<td>Male</td>
<td>0.0276</td>
<td>(0.0056)</td>
<td>0.0291</td>
<td>(0.0096)</td>
</tr>
<tr>
<td>log(Age^2)</td>
<td>0.0118</td>
<td>(0.0060)</td>
<td>-0.0350</td>
<td>(0.0104)</td>
</tr>
<tr>
<td>log(Income)</td>
<td>0.0317</td>
<td>(0.0103)</td>
<td>0.3757</td>
<td>(0.0178)</td>
</tr>
<tr>
<td>Student</td>
<td>0.00004</td>
<td>(0.0051)</td>
<td>-0.0093</td>
<td>(0.0087)</td>
</tr>
<tr>
<td>Married</td>
<td>-0.0940</td>
<td>(0.0090)</td>
<td>-0.0208</td>
<td>(0.0155)</td>
</tr>
<tr>
<td>Married</td>
<td>0.0180</td>
<td>(0.0076)</td>
<td>-0.0975</td>
<td>(0.0132)</td>
</tr>
<tr>
<td>Seoul</td>
<td>0.0020</td>
<td>(0.0098)</td>
<td>0.1480</td>
<td>(0.0170)</td>
</tr>
<tr>
<td>Busan</td>
<td>-0.0152</td>
<td>(0.0110)</td>
<td>0.0424</td>
<td>(0.0190)</td>
</tr>
<tr>
<td>Daejeon</td>
<td>-0.0110</td>
<td>(0.0127)</td>
<td>0.0643</td>
<td>(0.0219)</td>
</tr>
<tr>
<td>Const.</td>
<td>2.7014</td>
<td>(0.0665)</td>
<td>-1.8012</td>
<td>(0.1148)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.0191</td>
<td></td>
<td>0.1978</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td>49,500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: ML estimates of user demand function.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Interactions</th>
<th>Estimates</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Quantity</td>
<td>Mean</td>
<td>-0.0038</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>$D_{Male}$</td>
<td>0.000018</td>
<td>(0.000075)</td>
</tr>
<tr>
<td></td>
<td>User’s Age</td>
<td>0.000038</td>
<td>(0.000003)</td>
</tr>
<tr>
<td></td>
<td>Education Level</td>
<td>0.000153</td>
<td>(0.000017)</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.000879</td>
<td>(0.000046)</td>
</tr>
<tr>
<td>N. of Sections</td>
<td>Mean</td>
<td>0.0107</td>
<td>(0.0043)</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.0047</td>
<td>(0.0025)</td>
</tr>
<tr>
<td>$D_{telcom}$</td>
<td></td>
<td>1.1218</td>
<td>(0.2302)</td>
</tr>
<tr>
<td>Site’s Age</td>
<td></td>
<td>0.2272</td>
<td>(0.0387)</td>
</tr>
<tr>
<td>$D_{mail}$</td>
<td></td>
<td>-0.0498</td>
<td>(0.0759)</td>
</tr>
<tr>
<td>$D_{ecom}$</td>
<td></td>
<td>0.7340</td>
<td>(0.1763)</td>
</tr>
<tr>
<td>$D_{game}$</td>
<td></td>
<td>0.2747</td>
<td>(0.1666)</td>
</tr>
<tr>
<td>$D_{comm}$</td>
<td></td>
<td>0.2716</td>
<td>(0.2036)</td>
</tr>
<tr>
<td>$D_{mhpg}$</td>
<td></td>
<td>0.4935</td>
<td>(0.1661)</td>
</tr>
<tr>
<td>Constant for Outside Option</td>
<td></td>
<td>-5.9105</td>
<td>(2.1860)</td>
</tr>
</tbody>
</table>

Log-Likelihood                   -10800.68
Likelihood Ratio                  0.0720
N. of Observations               49,500

% of negative ad parameter       92.17 %
% of negative parameter for sections 1.1 %
Table 7: Estimated message processing rates ($g_{jt}$).

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naver.com</td>
<td>0.3612</td>
<td>0.4403</td>
<td>0.4390</td>
<td>0.5177</td>
</tr>
<tr>
<td>Daum.net</td>
<td>0.3875</td>
<td>0.4465</td>
<td>0.4516</td>
<td>0.4835</td>
</tr>
<tr>
<td>Nate.com</td>
<td>0.4360</td>
<td>0.5196</td>
<td>0.5201</td>
<td>0.6479</td>
</tr>
<tr>
<td>Yahoo.com/kr</td>
<td>0.4934</td>
<td>0.5441</td>
<td>0.5448</td>
<td>0.5932</td>
</tr>
<tr>
<td>Empas.com</td>
<td>0.4973</td>
<td>0.5344</td>
<td>0.5316</td>
<td>0.6448</td>
</tr>
<tr>
<td>Paran.com</td>
<td>0.3955</td>
<td>0.5370</td>
<td>0.5421</td>
<td>0.6651</td>
</tr>
</tbody>
</table>
Table 8: Elasticities with respect to advertising change.

<table>
<thead>
<tr>
<th></th>
<th>Price Elasticity, $\varepsilon_P$</th>
<th>Own Elasticity of $U_{jt}$, $\varepsilon_A^U$</th>
<th>Elasticity of $g_{jt}$, $\varepsilon_A^g$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congestion</td>
<td>Traditional</td>
<td></td>
</tr>
<tr>
<td>Naver.com</td>
<td>-0.7303</td>
<td>-0.6701</td>
<td>-0.0007</td>
</tr>
<tr>
<td>Daum.net</td>
<td>-0.7296</td>
<td>-0.6783</td>
<td>-0.0125</td>
</tr>
<tr>
<td>Nate.com</td>
<td>-0.7142</td>
<td>-0.6728</td>
<td>-0.0045</td>
</tr>
<tr>
<td>Yahoo.com/kr</td>
<td>-0.7002</td>
<td>-0.6734</td>
<td>-0.0054</td>
</tr>
<tr>
<td>Empas.com</td>
<td>-0.6974</td>
<td>-0.6696</td>
<td>-0.00002</td>
</tr>
<tr>
<td>Paran.com</td>
<td>-0.6907</td>
<td>-0.6698</td>
<td>-0.0002</td>
</tr>
<tr>
<td>Websites Pooled</td>
<td>-0.7104</td>
<td>-0.6723</td>
<td>-0.0039</td>
</tr>
</tbody>
</table>
Table 9: Estimated marginal costs.

<table>
<thead>
<tr>
<th>Website</th>
<th>Marginal cost(^a)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congestion</td>
<td>Traditional</td>
</tr>
<tr>
<td>Naver.com</td>
<td>2,567.30</td>
<td>7,369.90</td>
</tr>
<tr>
<td>Daum.net</td>
<td>3,399.53</td>
<td>7,004.65</td>
</tr>
<tr>
<td>Nate.com</td>
<td>5,065.03</td>
<td>9,216.09</td>
</tr>
<tr>
<td>Yahoo.com/kr</td>
<td>7,651.97</td>
<td>12,214.94</td>
</tr>
<tr>
<td>Empas.com</td>
<td>3,933.68</td>
<td>6,514.76</td>
</tr>
<tr>
<td>Paran.com</td>
<td>6,183.05</td>
<td>9,373.56</td>
</tr>
<tr>
<td>Websites Pooled</td>
<td>4,800.09</td>
<td>8,615.64</td>
</tr>
</tbody>
</table>

\(^a\) In Korean thousand won (\(\sim\) 1 US dollar).
Table 10: A difference in the process of drawing consumers’ attention by various advertising and media types.

<table>
<thead>
<tr>
<th>Media type</th>
<th>Targeted</th>
<th>Nuisance&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Stage 1 (Delivered)</th>
<th>Stage 2 (Decide to Give Attention)</th>
<th>Stage 3 (Action)</th>
<th>Stage 4 (Contact Advertisers)</th>
<th>Related Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telemarketing</td>
<td>Y</td>
<td>Y</td>
<td>Hear</td>
<td>Answer/Ignore</td>
<td></td>
<td></td>
<td>Choi et al. (2012)</td>
</tr>
<tr>
<td>Website banners</td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Y</td>
<td>Notice</td>
<td>Look/Ignore</td>
<td>Examine/Listen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search ads</td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N</td>
<td>Search</td>
<td>Read/Ignore</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Newspaper</td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Y</td>
<td>Notice</td>
<td>Read/Ignore</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Television</td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Y</td>
<td>Watch</td>
<td>Watch/Switch channel</td>
<td></td>
<td></td>
<td>Wilbur (2008)</td>
</tr>
<tr>
<td>Yellow Pages</td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N</td>
<td>Search</td>
<td></td>
<td></td>
<td></td>
<td>Rysman (2004)</td>
</tr>
<tr>
<td>Magazines</td>
<td>Y</td>
<td>N</td>
<td>Notice</td>
<td></td>
<td></td>
<td></td>
<td>Kaiser and Wright (2006)</td>
</tr>
</tbody>
</table>

<sup>a</sup> I consider only general type of media here. However, it is always possible for these media to target a certain group of people.

<sup>b</sup> This is the net effect of recipients’ attitude towards the ad messages. “Y” means that the recipients feel nuisance when they are exposed to signals of message delivery.