Estimating Online Advertising Demand Under Information Congestion

Dong Ook Choi*†

October 2013

UNFINISHED SCRIPT
– COMMENTS ARE WELCOME –

*Ph.D. Candidate, ESSEC Business School and THEMA, Université de Cergy-Pontoise, France
†This paper is a part of my thesis. I specially thank to my advisor Régis Renault. This paper has benefited from comments in the seminars and workshops at ESSEC Business School, Telecom Paristech, University of Virginia, and THEMA, Université de Cergy-Pontoise, EARIE 2013 conference at Universidade de Évora, and discussions with Simon Anderson, Federico Ciliberto, Olivier Donni, Nicolas Glady, Przemek Jeziorski, Nathalie Picard, and Thomas Trégouët. Financial supports from ESSEC Business School and THEMA are gratefully acknowledged.

1 Introduction

It is by now quite well-known that media that are financed through ad revenue face a double network externality problem: on one hand, the attractiveness of the media for its audience or readership depends on the intensity of the advertising activity that the media accepts; on the other hand, the attractiveness of the media for advertisers depends on the size of the audience or readership. In this research I study an additional externality that may have a substantial impact on the media platform's ability to deliver the customers' attention to advertisers and the interaction with the network externality described above. Receiver's attention is typically limited. If the quantity of ads on a platform is excessive, it is likely that different ad messages will be crowding out each other. Such a potential congestion effect has been pointed out and analyzed in some recent theoretical work (Van Zandt, 2004; Anderson and de Palma, 2009, 2012; Anderson et al., 2012).

The goal of this study is to empirically investigate the impact of the externality in the context of Internet platform competition. I suggest a structural model of online media platform given the heterogenous attention span of users. This attention model is fitted with our user-level data including website access records and demographic information.

I derive the advertising demand in a fashion similar to Rysman (2004). Using a theoretical framework that accounts for limited users' attention, it is possible to combine the demand model with the probability of users' message process. More specifically, I argue that I can explain the congestion effect by incorporating the message processing ratio directly into the amount of network effect from user side. The difficulty comes from the fact that the attention span of a 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.

I collect real market data in South Korea from 2004 to 2007. The data comes partly from Choi et al. (2012). I enriched this dataset with individual-level panel data. Internet websites in South Korea could be a good example because of highest internet penetration rate in the world during the study period. Besides the low advertising cost, the high accessibility for consumers and the fast growing market size make this industry appropriate in studying its economic property. Another merit comes from the fact that there was low degree of regulation on Internet advertising.

As a result, I estimate the probabilities of users' processing messages in each website. The range of probabilities is shown to be 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 higher market power, higher optimal ad level, lower 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.

Male, the young, and the less educated have lower attention span. Entry simulation shows that the network effect dominates the competition effect in the congestion model. These results can shed lights on policy implications in this industry.

This study is the first attempt to estimate the effect of information congestion on the basis of two sided market theory. In addition, thanks to the development in collecting data on Internet, I can utilize a unique and very detailed dataset which comprises both sides of the market. This paper is organized as follows: I go over previous studies in Section 2. Theoretical model on advertising demand with information congestion is presented in Section 3. I describe our dataset and estimation method in Section 4. Then, I provide the estimation results in Section 5 and conclude.

2 Previous Studies

Van Zandt (2004) and Anderson and de Palma (2009) focused on the effect of information congestion between senders and receivers without coordinating platforms. My work deals with this congestion problem in terms of media economics. Externality from information congestion may have been internalized to some degree by the existence of media platforms. As treating congestion problem with platform economics, I follow discussion from Anderson et al. (2012), but in an empirical way. This paper refers to Rysman (2004) as a basic two-sided market formulation and Goeree (2008) as an application of simulation methodology on limited information of consumers.

Although the importance of information congestion seems increasing, related studies are still relatively few in economics. Van Zandt (2004) and Anderson and de Palma (2009) are representative ones. Since physical communication cost lowers, advertising competition comes down to the human attention rather than physical communication channels. In accordance, these studies focus on the effect of limited human attention. Van Zandt (2004) presented a model which considers an effect of change in communication cost given a restricted capacity of receiver's attention span. He analyzed equilibrium within game theoretic framework in targeted communication. Anderson and de Palma (2009) also considered communication cost as an important factor, but they are different in that they introduced endogenous decision of receivers. That is, sender transmission and receiver examination work jointly (endogenously) to determine the provision of message and attention. One of strong assertions in both studies is that imposing tax for sending messages could help the market to increase social welfare by restricting excessive messages.

I indebted many parts to these studies such as the definitions of information congestion and its implications. However, my model tries to see the behavior of media platforms which control the number of messages to be sent. Media economics is not the main concern of these studies, even though Anderson and de Palma (2009) addressed in part. Therefore, the main difference of this study would be the endogenous decision of ad cost. Van Zandt (2004) and Anderson and de Palma (2009) assumed that the cost for sending message is exogenous, but the cost is endogenously decided by platforms in this study. Platforms exploit market power and try to maximize their profits by taking all the surpluses in the market. I demonstrate how much the level of information congestion would be in this settings with real market data.

My basic framework is based on two-sided market literature. Internet website is one of good examples of two-sided market. Network externalities from the other sides of market (user market and advertising market) characterize this market. Platforms make strategic decisions to have both markets on board. As seminal studies on this topic, Rochet and Tirole (2003), Caillaud and Jullien (2003), Armstrong (2006) examined how price structures would be changed with the different settings of platform competition. One of important results from these studies is that platforms subsidize consumer side market and impose higher price on advertising side market (known as "divide and conquer" strategy). This result works in my model in the sense that users usually don't pay for enjoying contents on websites. I develop demand functions from each side of the market and derive equilibrium condition following these studies.

Anderson et al. (2012) introduces interesting extensions of the traditional model of media economics: advertising congestion and multi-homing viewers. Each extension explains the puzzle that more entries would raise equilibrium ad level. I share the same motivation with the authors. However, they employed simple theoretical models with assuming exogenous viewer demand. My model will be different by considering endogenous viewer demand. Brown and Alexander (2005) empirically supports this result by showing that ad price rises by concentration. They explains that it is important to compare two elasticities (viewer's switching-off elasticity with respect to advertising and ad price elasticity with respect to advertising). However, they tested the relationship by a reduced-form model. The structural model could provide a complement explanation to this.

For the empirical study, Rysman (2004) is the representative one in the two-sided market and media economics studies. This study is the first one that suggested an empirical model of two-sided market in a media platform. It is an application on the Yellow Pages market in U.S and the study models their network effects in both sides of the market. Thus, more readership could lead to higher price-cost margin, and more ads have the same effect on readerships. For the basic framework, I follow this model in building two demand functions. However, my study differs in that I consider negative network externalities in user side market (nuisance effect) and information congestion.

Goeree (2008) is another important empirical study to be considered because of consumer's limited information in the model. The author examined PC market in U.S. within Berry et al.

(1995) (BLP henceforth) type framework, but she extended the model with heterogeneity of consumers having limited information. The idea of limited information is dealt with heterogeneous composition of choice sets among individuals. This is related to my model in the sense that I allow heterogeneity in individual attention span. As a result, there are high price-cost margin of PC industry. We follow this study especially in estimation strategy. However, my study will be different because I analyze from the perspective of media in which many different kinds of products are advertised, not from the perspective of a specific industry.

Table 1: A difference in the process of drawing consumers' attention by various advertising and media types.

Media type	Targeted	Nuisance[b]	Stage 1	Stage 2 (Decide to Give Attention)	Stage 3 (Active Examine)	Stage 4 (Action)	Related Study
Direct mails	Y	Y	Receive (Delivered)	Open/Throw-away	Examine	Contact Advertisers (make a phone call, click on a banner, and so on)	Anderson and de Palma (2009) and Van Zandt (2004)
Telemarketing	Y	Y	Hear (Phone ring)	Answer/Ignore	Listen		
Website banners	N[a]	Y	Notice	Look/Ignore			Choi et al. (2012)
Newspaper	N[a]	Y	Notice	Read/Ignore			*
Television	N[a]	Y	Watch	Watch/Switch channel			Wilbur (2008)
Yellow Pages	N[a]	N	Search		Look		Rysman (2004)
Magazines	Y	N	Notice		Read		Kaiser and Wright (2006)

[a] I consider only general type of media here. However, it is always possible for these media to target a certain group of people.
[b] 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.

3 Theoretical Model

3.1 Inverse demand for advertising slots

There are J websites that are indexed by $j \in [1, J]$ and the outside option (indexed by 0) in the market. 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 ads in j expecting that their messages can reach users in the platform j. The larger user base a platform has, the more expected profit for the advertisers. This will be captured as a positive network effect in the advertising demand function. I deal with the unsolicited advertising in the model. This implies that, *ex ante*, users don't know which message they would face until they access the website. Ads take spaces on the PC screen and distract users from watching contents so the number of ads can cause the nuisance effect in the user's utility function. This will be shown as a negative network effect in the user demand function. Facing these two demand functions, platforms coordinate advertisers and users to be 'on board.' I model this two-side market interactions: when a platform changes the ad quantity, the user demand will be affected by that and, in turn, the advertiser demand will be shifted by the change in the number of users. I assume also that there is no targeting or tracking technology.[1] This means that there is no matching between advertisers and consumers through the platforms.[2] Websites in this model are all independent players competing with each other (no ownership structure). The competition scheme is a static Cournot competition (quantity-setting game of the platforms). To specify the competition of websites, I introduce the assumption on the behavior of website users:

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

A1 is a traditional discrete choice assumption of viewers in media economics (e.g. Anderson and Coate (2005)). It says that when sending a message there is no way but the platform j to reach a user in j. This gives market powers to websites so they can act as monopolists in the advertising market ("competitive bottleneck" by Armstrong (2006)). Single-homing assumption can be restrictive for modeling the competition among Internet websites. However, this assumption simplifies the model to focus on the effect of limitation in the consumer's attention span. 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.

[1] Targeting and tracking are very interesting issues recently with Internet advertising (Athey and Gans, 2010; Athey et al., 2011). 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.

[2] Targeted advertising/justification in the data/counterfactual in the result. This can complicate the analysis of the effect from the limited attention span.

There is no restriction on the platform choice by advertisers. Advertisers can purchase ad slots in multiple websites. In doing this, advertisers need to know how much they can earn from the choice of the website. This should be dependent on how much ad space they buy and how many users (potential consumers) they can reach through the website. I follow Rysman (2004)'s formulation for modeling the advertising demand but modify the network effect considering the limited attention span. I assume that homogenous advertisers choose websites based on the expected number of users and other exogenous characteristics as well as the ad price. Since there is no targeting technology, the representative advertiser always chooses the website with the largest user base if all other conditions are equal. Different choices by advertisers can be explained by the website's exogenous characteristics and unobservables. The expected exposure for the representative advertiser is reflected in the look function which we denote as $L_{jt}(a_{jt}, \phi_{jt})$, where a_{jt} is the advertiser's ad quantity on platform j at period t and ϕ_{jt} is the expected number of views for an ad in platform j at period t. I introduce the second assumption to model advertiser's behavior:

Assumption 2. (small advertisers) *The quantity-setting platform j offers ad spaces to advertisers and the individual advertiser takes the aggregate ad quantity in j as exogenous. The ad quantity chosen by an individual advertiser doesn't affect the total ad quantity of the given website.*

This small advertiser assumption allows me to take aggregate ad quantity and number of users as exogenous to the individual advertiser. In other words, advertiser i in website j, the choice of a_{ijt} doesn't affect the aggregate number of ads in the platform A_{jt}. A1 and A2 tells us that websites are neither complements nor substitutes for individual advertisers. That is, any two look functions for different websites are independent (from the advertiser's viewpoint): $L_{jt} \perp L_{it}, \forall j \neq i$. This means that the cross-partial derivative of the profit by the representative advertiser is zero under A1 and A2.

The reason for this relationship is that advertising on a website targets totally different consumers from the other websites. Thus, an advertiser's choice of a in one platform does not affect the choice of a in other platforms. A1 also says that L_{ijt} is not a function of $\mathbf{A_{-jt}}$, implying that L_{ijt} is not directly affected by the number of ads in rival websites. 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.

In equilibrium, the representative advertiser sets the advertising demand at a point where expected profits from product sales equal the ad price. In other words, the advertiser makes a purchasing plan for ads based on the expected profit generated by potential consumers. Following Rysman (2004), we impose another assumption to construct a profit function:

Assumption 3. (proportionality) *the expected profit increases at a constant rate in the look* *function* L_{jt}.

By A3, I can construct a profit function for the representative advertiser and solve for the profit maximizing quantities of ads on the various websites. The total profit from the advertising mix is:

$$\Pi_t = \pi_{1t} L_{1t}(a_{1t}, \phi_{1t}) - P_{1t} a_{1t} + \cdots + \pi_{Jt} L_{Jt}(a_{Jt}, \phi_{Jt}) - P_{Jt} a_{Jt},$$

where Π_t is total profit across all the websites by the representative advertiser and P_{jt} is an average ad price in website j. π_{jt} is the advertiser's average profit per view in website j and we can separate this with the exposure amount L_{jt} by the proportionality assumption. The advertiser decides the number of ads to purchase, $\{a_{jt}\}_{j=1}^{J}$ to maximize its expected profit. The price paid for advertising on the website should be the marginal increase of the expected profit. We impose Cobb-Douglas functional form on L_{jt} so that $L_{jt}(a_{jt}, \phi_{jt}) = a_{jt}^{\alpha} \phi_{jt}^{\beta}$, where we expect that $\alpha > 0$ and $\beta > 0$ by the assumptions on the look function above. We solve for a_{jt} and derive the first order condition for website j. Then, we get the optimal ad level as,

$$a_{jt} = \begin{cases} \left(\dfrac{P_{jt}}{\alpha \pi_{jt} \phi_{jt}^{\beta}} \right)^{\frac{1}{\alpha - 1}}, & \text{if the advertiser chooses } j \\[2ex] 0, & \text{otherwise.} \end{cases}$$

By the assumption A2, each advertiser takes ϕ_{jt} as given although ϕ function depends on A_{jt} through information congestion (I discuss this more precisely in the next subsection). There exist corner solutions ($a_{jt} = 0\,, \exists j \in [1, ...J]$). The representative advertiser can decide not to choose a website if P is too high or ϕ is too low. This decision also depends on the exogenous characteristics (e.g. owned by a telecom company or providing differentiated services) and unobservable characteristics by researchers. π will capture the effect by these characteristics. 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, we get $A_{jt} = I_{jt} a_{jt} = \left(\frac{P_{jt}}{\alpha \bar{\pi}_{jt} \phi_{jt}^{\beta}} \right)^{\frac{1}{\alpha - 1}}$. Then, we can derive inverse demand function by solving for P_{jt},

$$P_{jt}(A_{jt}, \phi_{jt}) = \left(\alpha \bar{\pi}_{jt} \right) A_{jt}^{\alpha - 1} \phi_{jt}^{\beta}. \tag{3.1}$$

A_{jt} can be interpreted as supply side effect on the attention implying decreasing returns in ad spaces for the media j. ϕ_{jt} shows demand side effect meaning increasing returns in usage for the media j so that the demand for j would increase if advertisers expect more profits in j. Therefore, the sign of $(\alpha - 1)$ is expected to be negative and β be positive (we need to test this in the estimation).

An empirical form of advertising demand is derived by putting logarithms on both sides of Eq. (3.1).

$$\ln P_{jt} = \alpha^p \ln A_{jt} + \beta^p \ln \phi_{jt} + \mathbf{X_{jt}}\eta + \upsilon_{jt}, \tag{3.2}$$

where $\alpha^p = \alpha - 1$. Following the discussion above, we expect $-1 < \alpha^p < 0$ and $\beta^p > 0$ in the estimation. υ_{jt} is an error terms which captures unobservable characteristics. The effect by logarithm of $\alpha\bar{\pi}_{jt}$ will be explained by the observable characteristics ($\mathbf{X_{jt}}$) and unobservables (υ_{jt}). In the next subsection, the formulation of ϕ will be discussed where the effect of limited attention span comes in.

3.2 Information congestion and expected network effect from user side

The main concern of this study is to find the way how to estimate the effect of information congestion on the economic decisions of media platforms. I consider the limited capacity of users' attention that comes through the network effect ϕ_{jt} in the demand function. The expected network effect ϕ_{jt} shows the main difference between my model and the model in Rysman (2004). Even though Rysman (2004) deals with the advertising congestion in his model, he couldn't separate out the congestion effect from the effect of decreasing returns in willingness-to-pay. Specifically, Rysman (2004) has a_{ijt} and A_{jt} in $L(\cdot)$ function: the former is the willingness-to-pay of the advertiser and the latter is the business stealing effect among ad messages (i.e. information congestion so $\frac{\partial L_{jt}(\cdot)}{\partial A_{jt}} < 0$). I propose a different way to estimate the congestion effect given more detailed-level data on consumer usage. Instead of looking at the aggregate number of ads like in Rysman (2004), I model in more explicit way that a restriction is imposed on the network effect ϕ_{jt} in the advertising demand. Thus, A_{jt} is not shown in the $L(\cdot)$ function but appears in the ϕ_{jt} function in my model.

I consider display ads on websites which take an unsolicited way of advertising with no targeting.[3] Also, I am interested in general purpose platforms, not specialized ones. This is because the congestion of advertising messages would be more likely to occur in this environment so that I can focus on the congestion effect. I assume that advertisers expect to reach average recipients of the population and that users can't expect what messages they would face.

[3] 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.

Information congestion may have different definition by context. I consider "local congestion" in the sense that I deal with the effect of information congestion within a certain limitation. As localized congestion, I assume that information congestion is confined to one media type, Internet. This implies that effectiveness of Internet ads are independent of other types of media such as television, newspaper, magazine, or so. Another assumption about local congestion is that user's attention to ads doesn't get affected by contents on the same page. In other words, users are homogenous in attention on contents. This is quite strong assumption considering a widely accepted belief that a catchy content can help people recognize accompanying ads (George and Hogendorn, 2012). However, websites in our data (so called "Internet portals") normally have many different kinds of information in one page, so it's not easy to identify the effect of specific contents on ads. Therefore, I assume that platforms do not discriminate ads by accompanying contents. In sum, advertisers only care about the effect of Internet ads regardless of what contents are shown in the same page. I deal with this local congestion henceforce in this paper.

I postulate that the effect of limited attention comes in the expected number of users ϕ_{jt} as following:

$$\phi_{jt} = \phi_{jt}\left(g_{jt}, U_{jt}\right),$$

where U_{jt} is a number of users in website j and $g_{jt} = g\left(A_{jt}|U_{jt}\right) \in [0,1]$ represents a probability of accepting ad messages by given users. To generate the probability g_{jt}, I introduce an individual's attention span, m, which is unobservable to researchers. m is the maximum amount of messages (ads) processed by a user.[4] The unit of m is the same as A_j, the number of messages.

Assumption 4. (no externality) *Ad messages are identically and independently distributed. There is no externality in the distribution of ad messages and in the user's examining behavior. The number of visits in a website are uniformly distributed within the observation period and ad messages are also distributed in the same way. Therefore, m is i.i.d. across users.*

This assumption A4 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. Let $F(\cdot)$ and $f(\cdot)$ be a cumulative distribution and a density of m, respectively. Then, I consider "number of message losses out of A_{jt} given users in j at period t", $Loss(A_{jt}|U_{jt})$, as following:

$$Loss(A_{jt}|U_{jt}) = F(A_{jt}) \int_0^{A_{jt}} (A_{jt} - m) \frac{f(m)}{F(A_{jt})} \, \mathrm{d}m + (1 - F(A_{jt})) \times 0.$$

[4] g_{jt} can be interpreted as the examination function of individual receiver in Anderson and de Palma (2009).

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, $\frac{Loss(A_{jt}|U_{jt})}{A_{jt}} = \int_0^{A_{jt}} \left(1 - \frac{m}{A_{jt}}\right) f(m)\,\mathrm{d}m$, so the message processing rate, g_{jt} is:

$$
\begin{aligned}
g\left(A_{jt}|U_{jt}\right) &= 1 - \frac{Loss(A_{jt}|U_{jt})}{A_{jt}} = 1 - \int_0^{A_{jt}} f(m)\,\mathrm{d}m + \int_0^{A_{jt}} \frac{m}{A_{jt}} f(m)\,\mathrm{d}m \\
&= \underbrace{1 - F(A_{jt})}_{\text{the region where } m \geq A_{jt}} + \underbrace{\int_0^{A_{jt}} \frac{m}{A_{jt}} f(m)\,\mathrm{d}m}_{\text{the region where } m < A_{jt}} .
\end{aligned}
$$

Here, the formulation tells that each user would suffer information congestion when her attention span is less than the total ad messages in website j at time t (i.e. $m < A_{jt}$). I show that g_{jt} is differentiable with respect to A_{jt} in the Appendix. An important aspect of user behavior on examining display ads is that the average expectation on the remaining messages is the same as the *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 would accept min $\left(1, \frac{m}{A_{jt}}\right)$ of total messages. The resulting probability g_{jt} will be averaging these ratios over all users in website j at time t. g_{jt} also can be written $g_{jt} = \int \min\left(1, \frac{m}{A_{jt}}\right) \mathrm{d}F(m)$. By the assumption A4, the ϕ_{jt} function can be generated just by the multiplication between the user visits and the processing rate such as $g(A_{jt}|U_{jt}) \cdot U_{jt}$. Therefore, ϕ_{jt} can be written as

$$
\phi_{jt} = U_{jt} \int \min\left(1, \frac{m}{A_{jt}}\right) \mathrm{d}F(m). \tag{3.3}
$$

ϕ_{jt} is an expected outcome from the network effect: number of visits discounted by the actual ratio of taking ad messages into their cognition process. More specifically, Eq. (3.3) implies that expected profit of advertisers would be restricted if the attention levels of a portion of users are less than the total ads sent A_{jt}. Thus, advertisers would bear the ad cost even though there are some probable losses in the ad messages. In other words, the main idea in this model is that the reason for occurring information congestion is the heterogeneity of users in attention span distribution.

3.3 Attention level of each user

The sources of attention span are the user's inherent characteristics and the actual exposures in the website. A user's characteristics consist of her demographics (e.g. gender, age, and education level) and a unobservable fixed effect (e.g. inborn talents). Exposures can be measured by user behaviors (e.g. a duration of stay and a frequency of visits). In this subsection, I explain the generation of individual attention span more precisely.

Let m_{jkt} be the attention span of user k on website j at time t. Here, $k \in [1, \ldots, K]$ is an index for each user. m_{jkt} in the acceptance probability function g_{jt} is for incorporating the heterogeneity of individual attention into my model. I assume this variable to be positive and dependent on individual characteristics and behaviors. I assume that m follows log-normal distribution so $\ln m \backsim i.i.d. \mathcal{N}\left(\bar{m}, (\sigma^m)^2\right)$, where $\bar{m}$ is a mean and σ^m is a standard deviation of $\ln m$. Specifically, m_{jkt} function goes:

$$\ln m_{jkt} = \mathbf{D}_k' \alpha^m + \mathbf{H}_{jkt}' \beta^m + c^m + \sigma^m \zeta_{jkt}. \tag{3.4}$$

where $\mathbf{D_k}$ is a vector of user k's characteristics and $\mathbf{H}_{jkt}$ is a vector of user k's behavior in website j at time t. c^m is a constant. The error term ζ_{jkt} is in normal distribution with a standard deviation of σ^m. User behavior $\mathbf{H}_{jkt}$ is a (2×1) vector which represents user k's time spending in the website and the number of access (frequency of visits) during period t. I discuss how to formulate this vector in the estimation section. These indicators in $\mathbf{H}_{jkt}$ are necessary for me to capture the variation across websites in the attention span of a certain individual (demographic information remains the same across websites over time). For example, if a user stays longer and visits more frequent the website than others, she would probably have paid more attention to the website. Likewise, these indices can represent the intensity of a behavioral characteristic that a simple number of visits cannot explain. ζ_{jkt} captures the unobservable characteristics to researchers in attention level. One aspect of these unobservables is an inherent talent of the person which cannot be explained by the demographic information. I impose a distributional assumption on this error term: $\zeta_{jkt} \backsim i.i.d. \mathcal{N}(0,1)$. α^m, β^m, and σ^m are parameters to be estimated.

There are two important issues in formulating this attention level equation. The first issue is the unmeasurable property of consumer's attention span. No one can directly measure the size of m so I take an indirect way to generate m for each user. I simulate m's for all users given the distribution of ζ_{jkt} and Eq.(3.4). The second is the endogeneity issue with $\mathbf{H}_{jkt}$ in m function. The unobservable talent ζ_{jkt} can be correlated with the user behavior on a website. More specifically, the person with higher ζ could possibly stay longer or access more frequently on a website. I introduce additional equations to deal with this possible endogeneity. These two issues are discussed in more detailed way in the estimation section.

Note that the website characteristics do not directly affect the level of the individual attention span. I assume the separability between the formation of the attention span and the website characteristics. In other words, 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 A_{jt} and $\mathbf{X}_{jt}$ 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.

3.4 A model for user demand

The assumption A1 enables me to employ a discrete choice model. I apply the same strategy in estimating the user demand as in Nevo (2001) and Petrin (2002). Users choose a website based on the amount of ads and other exogenous characteristics. Price for accessing information is not charged but users are supposed to watch banners on the page. Subsidies are given to users because platforms can make profits by selling attention of users. There is the outside option where users can get information by accessing to other online sources. I begin with the conditional indirect utility function which consists of the website and the user's characteristics. Due to the homogeneity assumption in attention on contents, I don't need to consider the effect of the attention level m_{jkt} in the user utility function. A user in website j are indexed by k. The utility function of user k when she access website j is:

$$u_{jkt} = \rho_k A_{jt} + \mathbf{X_{jt}} \lambda_\mathbf{k} + \xi_{jt} + \epsilon_{jkt}, \tag{3.5}$$

where A_{jt} is the total number of ads on website j. $\mathbf{X_j}$ is a vector of exogenous characteristics of website j. ξ_{jt} is a unobserved error term with i.i.d. normal distribution which varies by websites and time periods. ϵ_{jkt} is an error term of user k in j and is assumed to follow i.i.d. logit distribution. Eq. (3.5) implies that the user utility for accessing website j is decided by the amount of ads, exogenous characteristics, and exogenous shocks. I employ the random coefficient model so ρ_k and λ_k are taste parameters as follows:

$$\begin{pmatrix} \rho_k \\ \lambda_k \end{pmatrix} = \begin{pmatrix} \rho \\ \lambda \end{pmatrix} + \mathbf{\Omega D_k} + \mathbf{\Sigma} \nu_\mathbf{k}, \tag{3.6}$$

D_k is a vector of user k's demographics. ρ and λ are mean tastes for A_j and $\mathbf{X_j}$, respectively. Ω is a parameter matrix for measuring the taste variation of demographics with respect to A_j and $\mathbf{X_j}$. ν_k is a fraction of heterogenous tastes following i.i.d. normal distribution. Σ is a scaling matrix. By assuming utility maximization, I define the set of variables to choose j as $V_{jt} \equiv \{(\mathbf{D_k}, \nu_k, \epsilon_{\cdot kt}) \mid u_{jkt} > u_{lkt}, \forall l \neq j\}$. Then, the market share of website j is

$$s_{jt} = \int_{V_{jt}} dG(\mathbf{D_k}, \nu_k, \epsilon_{jkt}),$$

where $G(\cdot)$ represents the joint distribution function of D_k, ν_k, and ϵ_{jkt}.

3.5 Equilibrium

By the assumption A1, the Internet advertising platform exploits a market power for the advertisers by providing differentiated products – in this case, exclusive number of users. Websites are in oligopolistic competition in the model. Here, I find the equilibrium ad quantity given limited attention span of users. I derive the equilibrium condition and estimate marginal costs as in Nevo (2001), Wilbur (2008), and Choi et al. (2012). Price-cost margin is derived and I find the factors that affect the market power of platforms.

There are many discussions on the characteristics of competition in media market. Previous empirical studies focus on quantity-setting game (see Rysman (2004) and Wilbur (2008) for the detailed discussion). Anderson and Coate (2005) show that there is no difference in the results between the price-setting and quantity-setting of media firms that are adversing-financed platforms. I also see the equilibrium scheme in the online banner advertising market as quantity-setting Nash equilibrium. My data shows that in the short run websites control the ad quantities and the menu prices rarely change (usually change in every six months to one year).

The timing of actions for the competition among websites is:

1. Websites decide how many sections for contents to provide. This is similar to the program choice in the television networks (Wilbur, 2008).
2. Websites design the composition of their displays in each section. Here, they decide the ratio between ads and contents in a page.
3. Websites set ad quantities, $\{A_{jt}\}_{j=1}^{J}$.

I consider only the constant marginal cost c_{jt} for website j at period t. The problem for the platform j at t is:

$$\max_{A_{jt}} \Pi_{jt}^{P} = P\left(A_{jt}, \phi_{jt}\right) A_{jt} - A_{jt} c_{jt}.$$

By applying first order condition with respect to A_{jt}, then we get

$$\frac{\partial \Pi^P_{jt}}{\partial A_{jt}} = P_{jt} - c_{jt} + \frac{\partial P_{jt}}{\partial A_{jt}} A_{jt} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial A_{jt}} A_{jt}. \tag{3.7}$$

The third term comes from the direct price change by A_{jt} and the fourth term shows the change in the network effect from user side. The expected network effect is $\phi_{jt} = \phi\left(g(A_{jt}|U_{jt}), U(A_{jt}, A_{-jt})\right)$ so the partial change of ϕ by A_{jt} 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. (3.7), then I get:

$$\frac{\partial \Pi^P_{jt}}{\partial A_{jt}} = P_{jt} - c_{jt} + \frac{\partial P_{jt}}{\partial A_{jt}} A_{jt} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \left(\frac{\partial \phi_{jt}}{\partial g_{jt}} \frac{\partial g_{jt}}{\partial A_{jt}} + \frac{\partial \phi_{jt}}{\partial U_{jt}} \frac{\partial U_{jt}}{\partial A_{jt}} \right) A_{jt}.$$

In equilibrium, $\frac{\partial \Pi^P_{jt}}{\partial A_{jt}}$ is set to zero so I can get the equilibrium ad quantity like follows:

$$A^*_{jt} = - \frac{P^*_{jt} - c_{jt}}{\frac{\partial P_{jt}}{\partial A_{jt}} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial g_{jt}} \frac{\partial g_{jt}}{\partial A_{jt}} + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial U_{jt}} \frac{\partial U_{jt}}{\partial A_{jt}}}, \tag{3.8}$$

where P^*_{jt} and other partial derivatives are the realizations at the equilibrium quantity A^*_{jt} so for example, $P^*_{jt} = P(A^*_{jt}, \phi(g(A^*_{jt}|U_{jt}), U(A^*_{jt}, A_{-jt})))$.

I compute the price-cost margin from Eq. (3.8) as follows:

$$\begin{aligned} \frac{P_{jt} - c_{jt}}{P_{jt}} &= -\frac{\partial P_{jt}}{\partial A_{jt}} \frac{A_{jt}}{P_{jt}} - \left(\frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial g_{jt}} \frac{\partial g_{jt}}{\partial A_{jt}} \right) \frac{A_{jt}}{P_{jt}} - \left(\frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial U_{jt}} \frac{\partial U_{jt}}{\partial A_{jt}} \right) \frac{A_{jt}}{P_{jt}} \\ &= -\left(\frac{\partial P_{jt}}{\partial A_{jt}} \frac{A_{jt}}{P_{jt}} \right) - \left(\frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\phi_{jt}}{P_{jt}} \right) \left(\frac{\partial \phi_{jt}}{\partial g_{jt}} \frac{g_{jt}}{\phi_{jt}} \right) \left(\frac{\partial g_{jt}}{\partial A_{jt}} \frac{A_{jt}}{g_{jt}} \right) \\ &\quad - \left(\frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\phi_{jt}}{P_{jt}} \right) \left(\frac{\partial \phi_{jt}}{\partial U_{jt}} \frac{U_{jt}}{\phi_{jt}} \right) \left(\frac{\partial U_{jt}}{\partial A_{jt}} \frac{A_{jt}}{U_{jt}} \right) \\ &= -\left[\varepsilon^P_A|_{g=\bar{g}\ and\ U=\bar{U}} + \varepsilon^P_\phi \cdot \varepsilon^g_A + \varepsilon^P_\phi \cdot \varepsilon^U_A \right], \end{aligned} \tag{3.9}$$

where $\varepsilon^P_A|_{g=\bar{g}\ and\ U=\bar{U}} = \frac{d \ln P_{jt}}{d \ln A_{jt}}|_{g=\bar{g}\ and\ U=\bar{U}} < 0$ is the elasticity of the price with respect to the ad quantity under the constant g_{jt} and U_{jt}, $\varepsilon^P_A = \frac{d \ln P_{jt}}{d \ln \phi_{jt}} > 0$ is the elasticity of the price with respect to the expected network effect, $\varepsilon^g_A = \frac{d \ln g_{jt}}{d \ln A_{jt}} < 0$ is the elasticity of message processing rate with respect to the ad quantity (see the Appendix for more details), and $\varepsilon_A^U = \frac{d\ln U_{jt}}{d\ln A_{jt}} < 0$ is the elasticity of the user demand with respect to the ad quantity. Eq. (3.9) 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. Thus, the bigger values of these elasticities of a firm could lead to the higher market power. Here, $\frac{\partial U_{jt}}{\partial A_{jt}} = \frac{\partial s_{jt}}{\partial A_{jt}} U^T$, where U^T is total number of users (market potential) which I could get from the user demand estimation.[5]

3.6 A social planner's problem

A social planner would choose a set of ad levels, $\{A_{jt}\}_{j=1}^J$, to maximize the sum of advertiser surpluses,

$$W = \sum_{j=1}^{J} \left[\int_0^{A_{jt}} P_{jt} \left(x, \phi \left(g(x|U_{jt}), U(x, \mathbf{A_{-jt}}) \right) \right) - c_{jt} A_{jt} dx \right].$$

When I take the first order condition:

$$\begin{aligned} \frac{\partial W_t}{\partial A_{jt}} &= P_{jt}\left(A_{jt}, \phi_{jt}\right) + \sum_{l=1}^{J} \left[\int_0^{A_{lt}} \frac{\partial P_{lt}\left(x, \phi_{lt}\right)}{\partial \phi_{lt}} \frac{\partial \phi_{lt}}{\partial A_{jt}} dx \right] - c_{jt} \\ &= P_{jt}\left(A_{jt}, \phi_{jt}\right) + \frac{\partial P_{jt}}{\partial \phi_{jt}} \frac{\partial \phi_{jt}}{\partial g_{jt}} \frac{\partial g_{jt}}{\partial A_{jt}} + \sum_{l=1}^{J} \left[\int_0^{A_{lt}} \frac{\partial P_{lt}\left(x, \phi_{lt}\right)}{\partial \phi_{lt}} \frac{\partial \phi_{lt}}{\partial U_{lt}} \frac{\partial U_{lt}}{\partial A_{jt}} dx \right] - c_{jt} \end{aligned} \tag{3.10}$$

Let A_{jt}^* be equilibrium level of ads and A_{jt}^O be socially optimal level of ads. A_{jt}^* and A_{jt}^O could be different if market is not in the perfect competition. A social planner should concern interactions among platforms. Even though $\mathbf{A}_{-jt}$ doesn't change directly P_{jt}, this would change U_{jt}. For the social planner, platforms can be complements or substitutes with each other depending on the sign of the cross partial derivative $\frac{\partial^2 W_t}{\partial A_{jt} \partial A_{kt}}, \forall k \neq j$.

[5] If s_{jt} is decided by discrete choice of users without considering random tastes, $\frac{\partial s_{jt}}{\partial A_{jt}} U^T = \rho s_{jt}(1 - s_{jt}) U^T$.

4 Data and Estimation

4.1 Market description

I look at so-called Internet search engines for the empirical analysis. Internet websites (including search engines) are suitable to study the 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 Korean population have accessed one of the websites in my data 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 to show the messages. Also, there is no targeting or tracking technology used in the websites during the study period. Therefore, consumers would ignore significant portions of the messages sent to them. Another possible reason could be that the cost per impression is lower than other types of media. All these characteristics could encourage advertisers to send messages excessively so it is highly probable for users to suffer the information overloads.

South Korea would be a good subject for this study because it is one of the highest and fastest Internet penetration countries. I select six biggest Internet search engines in Korea: "Naver.com," "Daum.net," "Nate.com," "Yahoo.com/kr," "Empas.com," and "Paran.com."[6] Almost 95% of Internet users in Korea visit at least one of these websites more than once in a month. It is interesting that Korean search engines provide not only searching function but also various contents such as news, blogs, advertisements, and so on.[7] These contents attract users to visit their website and make them to pay attention to ads. Contents are basically for free. These websites are, therefore, typical example of two-sided market where platforms can control the price structure in both markets.

Yahoo Korea is 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 biggest subscribers 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 interested in this market. In 2001, Nate.com is launched by SK telecom, the biggest 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

[6] 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.

[7] People often call this type of websites as an 'Internet portal' (Choi et al., 2012).

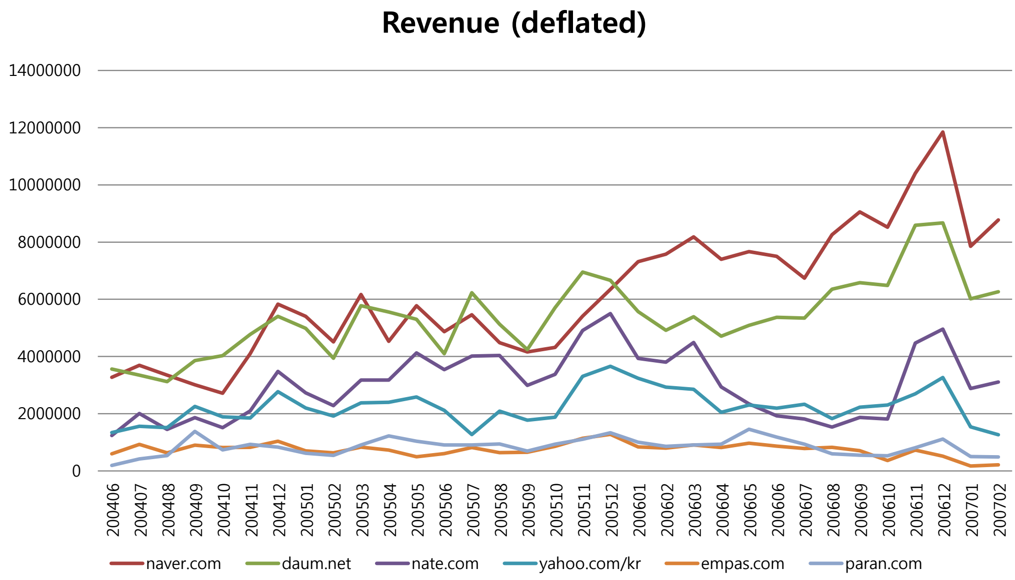

Figure 1: Revenue Growths of Banner Ads in Six Search Engines.

biggest 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.

4.2 Data

My dataset is composed of two parts: advertising and user market data. They are also generated in various precision from aggregate level to individual level. I discuss each of them in this subsection.

Advertising and website information in aggregate level

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 covering from June 2004 to February 2007 and so 198 observations in total (6 websites x 33 months). Descriptive statistics of advertising data is shown

Table 2: Description of Aggregate-Level Variables (Monthly Averages).

	Naver	Daum	Nate	Yahoo	Empas	Paran	Pooled[a]
Ad Price (in dollars)	16,022.39	15,728.24	25,081.58	30,025.09	9,323.36	15,308.55	18,581.54
Ad Quantity	633.91	572.39	334.94	205.21	95.91	114.39	326.13
N. of Advertisers	148.242	101.758	60.485	41.485	42.788	27.939	70.449
N. of Sections	18.697	22.424	25.030	20.879	12.303	11.061	18.399
Contract Periods	6.758	5.909	7.848	6.273	6.909	7.242	6.823
N. of Users	26,840,818	26,180,139	23,630,778	20,316,316	13,100,898	14,590,839	20,776,631
ARPU[b](in dollars)	0.235	0.211	0.130	0.113	0.058	0.068	0.136
Obs.				33			198

[a] Results from pooling all observations.
[b] Average Revenue Per User.

in Table 2.

I have monthly information such as ad revenue and the number of sections in each website. I also have the individual-level information, but I aggregate them to generate the ad price. I have advertiser's names, industry classifications, product names, ad contents, contract periods, sizes, and estimated prices in individual 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 user market. I put a time tag (year and month) on each observation. By doing so, I have monthly aggregate information. The individual price information in the data is not the real contract price but the estimated one by the surveying company. I would need more reliable price information in aggregate level. To generate the aggregate price, first I count the contract numbers by different sizes (i.e. different types) of ads in each time period. Then, I compute the price-adjusted quantity as follows:

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

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 estimated price of type i, and q_{ijt} is the number of ads of type i. $\frac{p_{ijt}}{p_{1jt}} q_{ijt}$ means the quantity adjusted by the relative price of type i to the basis price (p_{1jt}). The estimated price of the main banner in the top page of each website is assumed to be the basis price p_{1jt}. The monthly price is computed by dividing the revenue of website j in period t by this aggregated quantity Q_{jt}.

The number of sections approximates the amount of contents in a website.[8] I assume that the

[8]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 by its unlimited size unlike 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

decision of contents (and sections) would be prior to the decision of ad placement, so that number of sections is orthogonal to advertising demand.

I use starting/finishing day of each ad and generate "contract period" for each individual ad. Then, I aggregate the variable into website/month level. In Table 2, I present average periods according to websites. Nate.com has the longest average period, 7.85 days while Daum.net has the shortest one, 5.9 days. This variable of average contract periods is used as an instruments for the estimation.

In order to keep the assumption A1, I take the visits in a different website as the ones being made by different users. This could be counter-intuitive assumption and possibly bias the reality. However, allowing multi-homing behavior in user side could make the model extremely complicated in the two-sided market setting. To focus on the objective of estimating information congestion, I keep this single-homing assumption in this paper.[9] In Table 2, one can check that two biggest websites in the user market are Naver.com and Daum.net having more than 26 millions of user visits. Empas.com and Paran.com are the smallest firms with about one half of visits by Naver.com.

Six dummy variables are generated according to the websites' observable characteristics. *telcom* represents if the website is owned by telecommunication company. Nate.com and Paran.com are owned by SK telecom and KT, respectively, that are two biggest mobile companies in South Korea. *lmail* is one if the website provides large capacity (more than one giga byte) e-mail service. Three of six websites provide *lmail* service and two of them launched this service during the study period. *ecom* is one if the website provides e-commerce service. Three websites owned this service during the study period. *game* and *comm* is about offering gaming services and community services respectively. *comm* service is a grouping service that allows exclusive networking among members. *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.

Website usage data in individual level

For the user side market data, I collect visiting information for the six search engines during the study period. This data is gathered by Nielsen Korea. The Nielsen company runs a panel group for this survey. Each panel has a specialized application running in her/his web browser so that the company can track where, when, and how long she/he accesses to a certain website. This data can tell me about individual choice of the website as well as demographic information such as gender, age, education, income, and so on. Table 3 provides what information I have

user visits (see Choi et al. (2012)).

[9] I will tackle the multi-homing user model in the separate paper.

Table 3: Website usage and demographics (individual level, monthly averages).

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

1,157,246 obs. in total.

in individual level. Observations are in individual/website/month level. They are 1,127,546 observations in total. I simulate attention level of each user with this individual-level information. I also construct the user demand for a website with random coefficient model where random taste parameters are derived by interacting website characteristics with demographic information of users.

Page view (PV) is a count (number of hits) that increases every time user clicks on pages belong to a certain website. Average PV is highest in Nate.com and lowest in Empas.com during the study period (see Table 3). In some way, PV represents a performance in user market, but it depends on the structure of website. That is, if a website requires many clicks by construction, PV must be high on that website. In practice, PV is not used solely when evaluating a website. Duration means how long a user stays in a website. One index of a user's royalty on a website can be estimated by dividing duration by PV. For example, a user clicks on pages in Naver.com by 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 3). Daily frequency (DF) is a count of days in a month that a user accessed to a website. For example, users visited Nate.com 12.18 days in a month on average in Table 3. DF can be a good measure for user visits because PV is somewhat misleading for the reason explained above.

Finally, I have two interesting indices for measuring how royal a user is for a website: duration per PV and DF. Again, the former represents how long a user stay once she/he accessed to a website and the latter does how frequent she/he visits in a month. I show the distribution of these two indices in Figure 3. I made a vector using these two measures for describing the user behavior. I believe that this measure represents well the user's behavior on a website.

Nielsen also provides demographic information of each panel member. I utilize four variables explaining the behavior in this study: gender, age, education level, and income. Other demographics information such as jobs and regions is utilized as excluded instruments. As shown in Table 3, there are almost no differences in three variables among six websites. Average age and education level of users in six websites show less than 1 year difference and average income does about 50 US dollars difference between the lowest and the highest. This similarity among websites implies that these websites are very general purpose sites. In other words, it is not likely to think that these websites target a certain group of consumers; websites in our study deal with most of topics that fit for almost all kinds of people just as public television stations.

4.3 Estimation

Parameters are estimated by generalized methods of moment (GMM) procedure. I show how to derive moment conditions from the model in the following subsections. Three sets of moments are applied to estimate the advertising demand. Another set of moments following Berry et al. (1995) is separately applied for the user demand.[10]

Advertising demand estimation

I derive a residual for the moment condition of ad demand from Eq. (3.2) like follows,

$$\upsilon_{jt} = \ln P_{jt} - \alpha^p \ln A_{jt} - \beta^p \ln \phi_{jt} - \mathbf{X_{jt}}\eta. \tag{4.1}$$

In exogenous characteristics vector $\mathbf{X_{jt}}$, I put number of sections, six dummies of website characteristics (*telcom*, *lmail*, *ecom*, *game*, *comm*, and *mhpg*), age of website, age squared, and constant. To compute the residual υ_{jt}, I need the value of expected number of users ϕ_{jt} and also, the individual's attention level m_{jkt} that is unobservable by researcher. I simulate "hypothetical" attention span of users based on a certain distribution assumption of unobservables. Specifically, I generate m_{jkt} with exogenous characteristics, endogenous user behavior vector $\mathbf{H}_{jkt}$, and draws of random shocks ζ_{jkt} given parameter values (see Eq. (3.4)). I assume that ζ_{jkt} is in standard normal distribution so $\zeta_{jkt} \backsim i.i.d.\ \mathcal{N}(0,1)$.

Here, the user behavior vector $\mathbf{H}_{jkt}$ has two indices: the log of duration time per visit ($\ln dur_{jkt}$) and the log of the frequency of daily visits ($\ln f_{jkt}$). These two indices are potentially endogenous in m_{jkt} because the unobserved talents ζ_{jkt} can be correlated with them. The inherent ability of the message process (i.e. the higher ζ_{jkt}) could lead to longer duration or more frequent visits, or the other way around. I posit an additional equation about user behavior $\mathbf{H}_{jkt}$ as follows:

[10] I separated estimations of two demands because there is no cross-equation restrictions between advertising and user demand as Wilbur (2008).

$$\mathbf{H}_{jkt} = \begin{pmatrix} \mathbf{D}'_k\alpha^d + \beta_1^d A_{jt} + \mathbf{X}_{jt}^{h'}\beta_2^d + c^d + \epsilon_{jkt}^d \\ \\ \mathbf{D}'_k\alpha^f + \beta_1^f A_{jt} + \mathbf{X}_{jt}^{h'}\beta_2^f + c^f + \epsilon_{jkt}^f \end{pmatrix} \tag{4.2}$$

where $\mathbf{X}_{jt}^h$ is a subvector of exogenous characteristics of website j at time period t. ϵ_{jkt}^d and ϵ_{jkt}^f are the unobservable error terms which affects user k's behavior. α^d, β^d, c^d, α^f, β^f, and c^f are parameters to be estimated. Eq. (4.2) implies that an outcome of user behavior on website j depends on individual characteristics, ad quantity, and website characteristics (including contents). $\mathbf{H}_{jkt}$ could be higher with favorable website structure (related to $\mathbf{X_{jt}}$) so that it can draw more attention from users. Younger users (related to $\mathbf{D_k}$) might put higher values on a website than older users do.

My estimation strategy to deal with this potential endogeneity is that I regress $\mathbf{H}_{jkt}$ separately on related variables and I put the fitted residuals as explanatory variables in m_{jkt}. This can ensure that the parameter estimates of $\mathbf{H}_{jkt}$ is consistent. Then, the attention span m_{jkt} is:

$$\ln m_{jkt} = \mathbf{D}'_k\alpha^m + \mathbf{H}'_{jkt}\beta^m + c^m + \hat{\mathbf{E}}_{jkt}^{h'}\sigma^h + \sigma^m\zeta_{jkt}, \tag{4.3}$$

where $\hat{\mathbf{E}}_{jkt}^h\sigma^h = \sigma^d\epsilon_{jkt}^d + \sigma^f\epsilon_{jkt}^f$ is a fitted value of residuals from the estimation Eq. (4.2). $\hat{\mathbf{E}}_{jkt}^{h'}\sigma^h + \sigma^m\zeta_{jkt}$ is, therefore, the collective error term of m_{jkt} function conditional on ϵ_{jkt}^d and ϵ_{jkt}^f with following correlation matrix:

$$\begin{bmatrix} \zeta_{jkt} \\ \epsilon_{jkt}^d \\ \epsilon_{jkt}^f \end{bmatrix} \backsim i.i.d. \left(0, \begin{pmatrix} (\sigma^m)^2 & \sigma^d & \sigma^f \\ \sigma^d & 1 & 0 \\ \sigma^f & 0 & 1 \end{pmatrix} \right).$$

In this way, I take m_{jkt} as (hypothetical) attention span of individual panel members. I, here, know whether each individual has an attention level enough to receive the whole messages she faces. If m_{jkt} is larger than A_{jt}, I discard excessive amount. By doing so, I can compute ϕ_{jt} by averaging these acceptance ratios of each panel and applying it to Eq. (3.3).

I put the gender, $male_k$, the age of a user, age_k, and the square of the age, and the education level, edu_k in $\mathbf{D}_k$. By age, I can show if younger users tend to accept more information than older ones do. By education level, I can see if users with higher education level would not be curious about teasing advertisements and give less attention to them. $\mathbf{H}_{jkt}$ shows if a user visits more frequent and stays longer in each visit, her/his attention level would be higher. In other words, user behavior $\mathbf{H}_{jkt}$ implies the possibility to be exposed to ads would be increased.

I put gender, age, education level, and income in user behavior equation. I also put the number of sections for $\mathbf{X}_{jt}^{h}$ in user behavior function. From the estimation, I can find out how age and educational background affect a user's behavior of visiting and staying on the website. I believe that increasing ad quantity would decrease $\mathbf{H}_{jkt}$ and increasing contents would increase $\mathbf{H}_{jkt}$.

Parameters α^p, β^p, η, α^m, β^m, c^m, σ^m, σ^d, and σ^f are chosen in order that they minimize the GMM objective function, $\Lambda_A' Z_A W_A^{-1} Z_A' \Lambda_A$, where $\Lambda_A = \upsilon$ 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 efficiency and speed of the estimation, I applied two techniques in the process: importance sampling and antithetic acceleration. For constructing importance sampler, I draw samples more from the ones who might be under the information congestion problem. The probability is computed with the result from the first round estimation which is done without any prior information. Antithetic acceleration is useful when constructing simulated moments in aggregate level from the individual samples. Detailed explanation about these techniques will be provided in the Appendix.

4.4 User demand estimation

I assume that distributions of $\mathbf{D_k}$, ν_k, and $\epsilon_{\cdot kt}$ are independent with each other. By the logit distribution assumption of ϵ_{jkt}, the market share of j is derived as:

$$s_{jt} = \frac{1}{ns} \sum_{k}^{ns} \frac{\exp\left[\chi_{jt} + \sum_l^L x_{jt}^l \left(\sigma_l \nu_k^l + \omega_{l1} D_{k1} + \ldots + \omega_{ld} D_{kd}\right)\right]}{1 + \sum_m^J \exp\left[\chi_{mt} + \sum_l^L x_{mt}^l \left(\sigma_l \nu_k^l + \omega_{l1} D_{k1} + \ldots + \omega_{ld} D_{kd}\right)\right]}, \tag{4.4}$$

where x_{jt}^l is l'th variable of $\mathbf{X_{jt}}$, and D_{kd} is d'th variable of $\mathbf{D_k}$. χ_{jt} is a mean utility function for website j such that

$$\chi_{jt} = \rho A_{jt} + \mathbf{X_{jt}} \lambda + \xi_{jt}. \tag{4.5}$$

I apply BLP-type contraction mapping to compute residuals for the moment condition of user demand. Contraction mapping is well-defined here to derive the mean utility level from observed market share (Berry et al., 1995; Nevo, 2001; Goeree, 2008). Let χ_{jt}^r be a mean utility of users in website j in r'th iteration:

$$\chi_{jt}^{r+1} = \chi_{jt}^r + \ln(s_{jt}) - \ln\left(s\left(A_{jt}, X_{jt}, \chi_{jt}^r | \theta\right)\right), r = 0, \ldots, R, \tag{4.6}$$

where $s_{jt}(\cdot)$ is market share of website j in user market and θ is a vector of parameters in user demand function, Eq.(3.5) and Eq. (3.6). χ_{jt}^R approximates to χ_{jt} after enough iterations. This is a well-established method, but it takes much time to converge because there is extra inner loop to compute $\hat{\chi}$ rather that the iteration for searching proper parameter values. Once I have the optimal mean utility $\hat{\chi}_{jt}$, then I can derive residual vector which is for building moment condition:

$$\xi_{jt} = \hat{\chi}_{jt} - \rho A_{jt} - \mathbf{X_{jt}}\lambda, \tag{4.7}$$

where I include the number of sections, six dummies of website characteristics (*telcom*, *lmail*, *ecom*, *game*, *comm*, and *mhpg*), age of website, age squared, and constant in $\mathbf{X_{jt}}$. We also estimate parameters ρ, λ, Ω, and Σ in Eq.(3.5) and Eq. (3.6) by GMM procedure. We minimize the GMM objective function, $\Lambda_U' Z_U W_U^{-1} Z_U' \Lambda_U$, where $\Lambda_U = \xi_{jt}$ is a vector of error terms from Eq. (4.7), Z_U is a vector of instruments for the moment condition. W_U is a weight matrix that is a consistent estimate of $E[Z_U' \Lambda_U \Lambda_U' Z_U]$ as in advertising demand estimation.

In user demand equation, A_{jt} is endogenously decided to the mean utility. This is obvious because the high level of the aggregate utility would increase the advertising demand (see Eq. (3.1)). I apply proper instruments to deal with this endogeneity. I include the average contract periods and average characteristics of rival firms (so called BLP instruments) as the excluded instruments. More detailed discussions and justifications will be provided in the identification section.

4.5 Individual-level user choice equation

*[TO BE INCLUDED]

4.6 Identification

Endogenous variables in inverse demand of advertising, Eq. (4.1), are A_{jt} and ϕ_{jt}. As my model says, the ad quantity and the expected users explains the ad price. Inversely, unobservable quality factors that change the ad price (e.g. server network capacity, the provision of popular contents, or so) would affect the ad quantity (or the ad demand) and the expected number of users. In user demand equation, A_{jt} is endogenously decided to the mean utility (see Eq. (4.7)). I deal with these endogeneity problems by applying proper instruments in GMM estimation.

On the choice of GMM instruments, I use exogenous shocks in rival firms as in Berry et al. (1995) and Nevo (2001). I apply Berry et al. (1995) approximation of a polynomial of exogenous

Table 4: First-stage estimation of ad demand equations.

Advertising Demand	$\log(AdQuantity)$		$\log(UserVisits)$	
	Coef.	Std. Err.	Coef.	Std. Err.
Avg. rivals' contract days	-0.2219	(0.1948)	0.0285	(0.0562)
Avg. contract days	-0.0246	(0.1098)	-0.0171	(0.0317)
Total internet users	0.00009	(0.00009)	0.00004	(0.00003)
$\log(Sections)$	0.6222	(0.3009)	0.2081	(0.0868)
Age	0.5619	(0.1699)	0.3519	(0.0490)
Age^2	-0.0259	(0.0126)	-0.0286	(0.0036)
D_{lmail}	-0.2096	(0.2049)	0.0677	(0.0591)
D_{ecom}	2.4626	(0.5948)	0.6061	(0.1716)
D_{game}	-1.1074	(0.5805)	0.0212	(0.1675)
D_{comm}	2.7037	(0.6365)	0.1610	(0.1837)
D_{mhpg}	-1.5133	(0.5688)	0.0301	(0.1642)
D_{telcom}	2.6785	(0.7691)	0.5733	(0.2220)
Const.	-3.8380	(2.8168)	13.3486	(0.8129)
Adj. R^2	0.5547		0.2168	
Obs.	198			
User Demand	$\log(AdQuantity)$			
	Coef.	Std. Err.		
Avg. rivals' sections	-0.0464	(0.0157)		
Avg. rivals' D_{lmail}	0.3723	(0.5223)		
Avg. rivals' Age	0.0036	(0.0984)		
Avg. rivals' duration	0.0038	(0.0023)		
Avg. rivals' contract days	-0.1833	(0.0839)		
Avg. contract days	0.0243	(0.0471)		
N. of sections	0.0303	(0.0077)		
D_{telcom}	0.1237	(0.3778)		
D_{lmail}	-0.5477	(0.1272)		
D_{ecom}	0.3515	(0.2609)		
D_{game}	0.2640	(0.2469)		
D_{comm}	0.5691	(0.3209)		
D_{mhpg}	0.1809	(0.2438)		
Age	0.1649	(0.0581)		
Const.	4.5879	(0.7832)		
Adj. R^2	0.7920			
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 in bold.

variables: The number of sections, *lmail*, age of websites, and the average contract periods. I put these variables all in averaged value of rival firms.[11]

[11] Chamberlain (1987) suggest that the optimal instruments for GMM would be "conditional expectation of the derivative of the disturbance vector wrt the parameters of interest, conditional on the set of exog. variables," that is, $E\left[\frac{\partial v(\theta_o)}{\partial \theta}|z\right]$ where θ_o is a vector of optimal parameters. Berry et al. (1995) used an approximation by a

Excluded instruments can be seen as in three groups: average behavior of advertisers, average characteristics of users, and overall market condition. As the advertiser's behavior I include average contract days. To use this variable as an instrument, I need to make an assumption that advertisers decide the contract period of their ads before they decide the ad quantity. Thus, change in the ad quantity doesn't impact the contract period. If an advertiser needed different ad messages (i.e. different brands, different ad contents, or so), they would have put more ads, not decreasing contract period. This is true for shorter period contracts, and I confirmed in the data that 70% of ads are less than 7 days period. This assumption makes contract periods exogenous to ad quantities. I choose some average profiles of users as 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.

Here, I need to mention the panel property of our data set. Panel structure of data in Nevo (2001) seems to be comparable to my case. In Nevo (2001), exogenous characteristics of ready-to-eat cereals should be constant with a certain brand. For example, the percentage of sugar, the weight of the product, and so on. To deal with the endogeneity, Nevo (2001) used average price of cereals in neighbor cities. However, being different from Nevo (2001), some of the exogenous characteristics in my data have variation over time. Continuos variables such as number of sections, age of websites, and total internet users change over time. Also, dummy variables such as *lmail* and *mhpg* are not constant during the period because these services are newly introduced in some of websites during the study period.

5 Result

5.1 Parameter estimates of the model

I consider two different specifications of the estimation. Model (I) is the traditional demand function without considering the limitation in the attention span. My model is shown in the specification (II). Estimation result of aggregate level variables are shown in Table 5. Signs of parameters of interest, α^p and β^p, are consistent with the theory. The signs of α^p in all specifications are shown to be negative as I expected. The parameter estimate 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 user demand, ϕ_{jt}, is significantly positive. This shows that the more expected views on a website raises the price, showing the network externality from user side.

polynomial in the relevant variables and Goeree (2008) derived a numerical approximation. I follow Berry et al. (1995) in this study.

Table 5: GMM estimates of the advertising demand function.

	(I) Traditional Model	(II) Congestion model		
	Ad Demand	Ad Demand	Attention Eq.	
Ad Quantity ($\hat{\alpha^p}$)	-0.7778	-0.6112	Male	-3.3172
	(0.1237)	(0.0008)		(0.1682)
Expected User Demand ($\hat{\beta^p}$)	0.6631	0.6553	Age	-3.3473
	(0.0405)	(0.0059)		(0.1403)
Num. of Sections	0.3618	0.4846	Age^2	1.4376
	(0.1902)	(0.0022)		(0.1923)
Age of Website	0.0114	-0.0738	Edu	4.2383
	(0.0243)	(0.0229)		(0.4556)
D_{lmail}	-0.4081	-0.4061	Dur/PV	-28.1084
	(0.0691)	(0.0014)		(0.0380)
D_{ecom}	03516	-0.0233	Frequency	-4.8696
	(0.1791)	(0.0075)		(0.0643)
D_{game}	0.6632	0.9111	Const.	98.4117
	(0.1265)	(0.000014)		(0.3712)
D_{comm}	-0.4743	-0.9768	σ^d	37.6916
	(0.3564)	(0.0150)		(0.1572)
D_{mhpg}	0.5518	0.7678	σ^f	17.1820
	(0.1640)	(0.0012)		(0.4437)
D_{telcom}	-0.1125	-0.6840	σ^m	24.9752
	(0.0996)	(0.0502)		(0.4736)
Const.	1.2473	1.8834		
	(0.3978)	(0.0021)		
GMM Objective	9.8693	1.4496		

Note: Standard errors are reported in parentheses.

Number of sections that approximates content amounts has positive effect on the ad demand. This tells us that advertisers value contents because ad spaces would rise with the number of sections. Also, having more sections can give a signal that the website has more capability to invest in contents and to draw users. Among dummy variables, parameters for *lmail*, *ecom* and *comm* have negative estimates and others have positive estimates. Negative effect of large e-mail service could imply its low contribution to advertising. This could also mean that websites could have not been successful in linking e-mail services with advertising. In the same way, e-commerce and community services might not effective for advertising delivery. Gaming services on websites are usually online services that connect players. *mhpg* are basically networking services like Facebook that encourage users to gather around on the platform. Therefore, *game* and *mhpg* have an important feature in common: they keep users in the website connected online. This feature might have produced good outcomes for advertisers. Estimates of *telcom* dummy are significant only in specification (II). *telcom* has estimated negatively to the demand. The reason might be that the websites belong to the large telecom companies are not the first movers but followers so they have a certain disadvantage in the market.

Table 6: The additional estimations of user behavior equations.

	Log (Duration per PV)		*Log (Frequency)*	
	Coef.	Std. Err.	Coef.	Std. Err.
log($AdQuantity$)	-0.0183	(0.0037)	-0.0207	(0.0063)
log($Sections$)	0.0867	(0.0149)	0.4217	(0.0258)
$WebsiteAge$	0.0644	(0.0080)	0.3009	(0.0137)
$WebsiteAge^2$	-0.0045	(0.0006)	-0.0245	(0.0011)
D_{lmail}	-0.0020	(0.0096)	-0.2932	(0.0167)
D_{ecom}	-0.0344	(0.0308)	0.8581	(0.0532)
D_{game}	0.0967	(0.0289)	0.0355	(0.0499)
D_{comm}	0.0777	(0.0334)	0.3961	(0.0577)
D_{mhpg}	-0.0463	(0.0286)	0.0404	(0.0494)
D_{telcom}	-0.0130	(0.0391)	0.6573	(0.0675)
$Male$	0.0276	(0.0056)	0.0291	(0.0096)
log(Age)	.	.	.	.
log(Age^2)	0.0118	(0.0060)	-0.0350	(0.0104)
log(Edu)	0.0317	(0.0103)	0.3757	(0.0178)
log($Income$)	0.00004	(0.0051)	-0.0093	(0.0087)
$Student$	-0.0940	(0.0090)	-0.0208	(0.0155)
$Marride$	0.0180	(0.0076)	-0.0975	(0.0132)
$Seoul$	0.0020	(0.0098)	0.1480	(0.0170)
$Busan$	-0.0152	(0.0110)	0.0424	(0.0190)
$Daejeon$	-0.0110	(0.0127)	0.0643	(0.0219)
Const.	2.7014	(0.0665)	-1.8012	(0.1148)
Adj. R^2	0.0191		0.1978	
Obs.	49500			

I present results of the attention level estimation in the last column of Table 5. All parameters are shown significant. According to the estimates, the parameter for male is shown negative to the attention span. This implies that male users have shorter attention span than females do. Parameter for age and age squared say that the attention span increases with age. People with higher education pay more attention to ad messages. The effect of duration and frequency on the attention level is significant and negative. This seems a bit surprising but the reason might be the inclusion of error terms of the additional user-behavior equations. The parameters for the user behaviors tells that the person with low attention level could have stayed longer and visited more frequent.

In order to produce the error terms ϵ^d and ϵ^f, I run the user behavior equation Eq. (4.2) first. The estimation results of Eq. (4.2) are shown in Table 6. The parameters for the ad quantity are negatively significant and have the same sign in *duration* and $frequency$. In contrast, content amount has a positive effect on both indices. Therefore, contents draw users but ads discourage them as I expected.

Table 7: GMM estimates of user demand function.

	Coef.	Std. Err.	Interactions: Sigma	Age	Edu
Ad Quantity ($\hat{\alpha}$)	-0.0053	(1.9350)	0.0026	-0.0087	0.0160
			(1.04e-5)	(3.1e-5)	(7.76e-5)
Num. of Sections	0.0568	(1.4730)	0.0008		-0.0068
			(3.76e-6)		(5.92e-5)
D_{telcom}	-2.3239	(0.2853)			
D_{lmail}	0.7099	(0.4322)			
D_{ecom}	-2.5745	(0.6725)			
D_{game}	2.3617	(2.7859)			
D_{comm}	-1.8127	(2.0267)			
D_{mhpg}	2.3516	(1.4730)			
Age of Website	0.0578	(0.3142)			
Age Squared	0.0034	(0.2783)			
Const.	-0.5223	(0.8647)			

Standard errors are reported in parentheses.

5.2 Result from user demand estimation

I present the result from user demand estimation in Table 7. Random coefficients are assigned only to ad quantity and number of sections. Also, I put zero restrictions on some of random coefficients for the estimation. The resulting parameter estimate for ad quantity is shown to be negative as expected, but not significant. The effect of number of sections is positive and significant. Estimates of three dummies are significant: *telcom*, *ecom*, and *mphg*. These results are quite consistent with the result from advertising demand. Advertisers seem to value the same characteristics of websites as users do by this result.

5.3 Message processing rates

One interesting estimate is the probability $g_{jt} = \int \min\left(1, \frac{m}{A_{jt}}\right) dF(m)$ in each model. This value is the result of averaging message processing rates for each user. I present the values of g_{jt} in Table 8. Estimated values are different by 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.[12]

[12]There is a potential bias in estimating g_{jt} due to the single-homing assumption of users. If users can visit multiple search engines, it could be possible that they can be exposed to ads on different search engines. It follows that these websites would compete for user attention. Therefore, estimating the degree of information congestion could be even more complicated.

Table 8: Estimated message processing rates (g_{jt}).

	Min	Mean	Median	Max
Naver.com	0.3612	0.4403	0.4390	0.5177
Daum.net	0.3875	0.4465	0.4516	0.4835
Nate.com	0.4360	0.5196	0.5201	0.6479
Yahoo.com/kr	0.4934	0.5441	0.5448	0.5932
Empas.com	0.4973	0.5344	0.5316	0.6448
Paran.com	0.3955	0.5370	0.5421	0.6651

5.4 Elasticities and demand functions

I present elasticities of prices with respect to ads in the Table 9. In other words, these values are the amount of price changes by the 1% change of ad level.[13] First two columns are the elasticities computed on the assumption that network effect from user market doesn't change by the small shift of ad level. The resulting elasticities, therefore, only account for the effect on ad price by the change in the ad level, but not in the number of users. First column is about elasticities computed under the assumption that information congestion model (specification (II) in the previous subsection) is the true one. In the first column, one can see that the values vary by websites. Even though there is no change in number of users, g_{jt} function changes where ad level is included. In the second column, there are elasticities when we assume that model (I) is the true model. As one can expect, all values are the same in the second column. It is obvious because parameter estimates are assumed to be equal across the websites.

Elasticities in the third and the fourth columns are computed allowing the change of network effect from user side. In the same way with the first and the second columns, third column is the result under the congestion assumption and the fourth is under the no congestion assumption. What I can find here is that the absolute values of elasticities in the third column are smaller than the ones in the fourth column. In the fifth column, I show the change in the number of users by 1% rising in ad level. Only in paran.com, number of users goes up. It is possible due to the random coefficient model specification of user demand function.

These elasticities are also interpreted as negative values of Lerner index, that is, a market power of platforms in the advertising market. Overall, numbers are smaller in the congestion case. This implies that the result could be biased in the market power estimation without considering congestion. Specifically, ignoring information congestion, I overestimate the amount of market powers. I can estimate marginal costs using these elasticities.[14] Estimated marginal costs are 1,896.2 USD in specification (II) and 999.1 USD in specification (I) when allowing U to change. This shows that if one doesn't consider the effect of information congestion, marginal costs might

[13] I compute these values numerically by assuming that platforms are in the average market.

[14] If I say e is an elasticity (the sum of all elasticities, see Eq. (3.9)), then the marginal cost mc can be computed by $p \times (1 + e)$.

Table 9: Price and user elasticities with respect to advertising change.

	No change in U		Change in U		Elasticity of $U(A)$
	Congestion	No congestion	Congestion	No congestion	
Naver.com	-0.6938	-0.7778	-0.7898	-0.8672	-0.1367
Daum.net	-0.6907	-0.7778	-0.7910	-0.8711	-0.1428
Nate.com	-0.6870	-0.7778	-0.7848	-0.8679	-0.1393
Yahoo.com/kr	-0.6909	-0.7778	-0.7543	-0.8342	-0.0905
Empas.com	-0.6871	-0.7778	-0.7067	-0.7890	-0.0280
Paran.com	-0.6874	-0.7778	-0.7130	-0.7959	0.0367
Websites Pooled	-0.6873	-0.7778	-0.7820	-0.8658	-0.1288

[a] In Korean million won ($\simeq$ 1,000 US dollars).

be underestimated.

I draw the inverse demand functions for each specification in Figure 4. The horizontal axis stands for the number of ads and the vertical axis stands for the ad price in US dollar. Using above estimates, optimal ad level is about 3,385 in specification (I) and about 2,207 in specification (II).[15] One can see that the optimal ad level can be overestimated without considering the information congestion. This difference depends on the slopes of two demand functions and the gap between two marginal costs.

Figure 5 shows the growth of average attention spans in the ad quantity.

5.5 Entry simulation

Table 10 shows the results from the platform's entry simulation. I assume the symmetric websites in the average market. In case of more than two platforms, a platform with the smallest ad quantity would take the whole user demand. Therefore, the necessary condition for the Nash equilibrium is the symmetry of the market share in the user market. Then, I solve for the equilibrium ad quantity in the Cournot-Nash equilibrium in the advertising market. I compare the results of specification (I) and (II). Both equilibrium ad prices and quantities are smaller in the traditional model so much smaller profits. Surprising results are the total advertising surpluses in two models. In the congestion model, total surpluses are decreasing in the number of entrants while in the traditional model, surpluses are increasing in the number of entrants. The network effects in the congestion model seem dominate the competition effect.

[15] All figures in the graphs are estimated numerically.

Table 10: Entry simulation results in two models.

Congestion Model

N. of Entrants	Ad Price	Equilibrium Ad Quantity	Optimal Ad Quantity	Profit	Total Surplus
1	21,057.77	906.29	29,867.23	19,082,542.86	176,544,751.2
2	16,002.33	667.18	14,773.52	10,674,555.23	172,809,006.2
3	13,653.28	556.08	9,771.59	7,590,392.29	170,382,235.8
4	12,211.25	487.87	7,283.00	5,955,657.15	168,559,599.9
5	11,206.17	440.34	5,798.53	4,932,592.03	167,103,650.8
6	10,452.66	404.70	4,811.64	4,228,272.09	165,871,767.6
7	9,858.57	376.60	4,110.10	3,710,831.4	164,827,250.3
8	9,372.87	353.63	3,585.81	3,312,601.42	163,893,991.2
9	8,966.28	334.40	3,179.08	2,996,390.03	163,077,518.4
10	8,619.48	317.99	2,854.14	2,739,041.49	162,319,326.9

Traditional Model

N. of Entrants	Ad Price	Equilibrium Ad Quantity	Optimal Ad Quantity	Profit	Total Surplus
1	16,799.42	789.81	29,744.92	13,267,406.67	261,972,427.4
2	13,075.53	603.67	16,472.27	7,892,262.34	263,907,563.3
3	11,304.57	515.14	11,657.82	5,822,445.06	264,875,130.1
4	10,201.24	459.99	9,122.08	4,691,458.20	265,487,714.1
5	9,423.55	421.11	7,541.68	3,967,392.11	265,898,968.9
6	8,834.61	391.67	6,455.92	3,459,290.39	266,199,698.7
7	8,367.08	368.30	5,660.81	3,080,629.64	266,430,512.1
8	7,983.34	349.12	5,051.67	2,786,160.91	266,610,428.2
9	7,660.43	332.98	4,569.01	2,549,778.09	266,763,034.8
10	7,383.42	319.13	4,176.46	2,355,297.55	266,864,751.5

5.6 Robustness Test

I employ different estimation methods to check the robustness of the estimations. The first one is the simultaneous estimation of additional user-behavior equations with the ad demand function. The result is shown in Table 11 and Table 12.

6 Conclusion

Website users are heterogenous in accepting advertising messages. I consider this heterogeneity as the model of information congestion in advertising demand. I follow Rysman (2004) in building traditional two-sided market structure, and extend the model with the specification of

Table 11: Robustness Check: GMM estimates of the advertising demand function.

	Ad Demand		Attention Eq.
Ad Quantity ($\hat{\alpha^p}$)	-0.6112	Male	-2.4011
	(0.0008)		(0.0005)
Expected User Demand ($\hat{\beta^p}$)	0.6553	Age	0.4978
	(0.0059)		(0.0013)
Num. of Sections	0.4846	Age^2	-0.5019
	(0.0022)		(0.0112)
Age of Website	-0.0738	Edu	-0.6282
	(0.0229)		(0.0018)
D_{lmail}	-0.4061	Dur/PV	0.8607
	(0.0014)		(0.0084)
D_{ecom}	-0.0233	Frequency	0.4975
	(0.0075)		(0.0005)
D_{game}	0.9111	Const.	2.4659
	(0.000014)		(0.0015)
D_{comm}	-0.9768	σ^m	9.5193
	(0.0150)		(0.0023)
D_{mhpg}	0.7678		
	(0.0012)		
D_{telcom}	-0.6840		
	(0.0502)		
Const.	1.8834		
	(0.0021)		
GMM Objective		18.2626	

Standard errors are reported in parentheses.

Table 12: Robustness Check: Parameter estimates in additional user behavior equations.

	(II) *duration*		(II) *frequency*	
	Coef.	Std. Err.	Coef.	Std. Err.
$Male$	0.7162	(0.0046)	-1.4084	(0.0062)
Age	1.2500	(0.0020)	3.3324	(0.0033)
Age^2	-0.0905	(0.0016)	-1.5167	(0.0136)
Edu	-1.8622	(0.0041)	-0.6785	(0.0147)
$Income$	-0.1996	(0.0012)	-1.0788	(0.0038)
Ad Quantity	-0.3961	(0.0015)	1.2857	(0.0013)
N. of Sections	1.0462	(0.0009)	0.1747	(0.0019)
Error Correlation $\hat{\sigma}$	3.8197	(0.0032)	2.1367	(0.0085)
Const.	9.7653	(0.0102)	3.8410	(0.0054)

Standard errors are reported in parentheses.

information congestion. I show that there is a possibility of bias in demand estimation without considering advertising congestion.

By the two-sided market property, 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 it is usual to make this assumption in conventional media economics, users can choose multiple websites in reality. The future research will be exploring the multihoming behavior by consumers in this market.

A Appendix: Proof of the differentiability of g_{jt}

I suppress the subscripts j and t for the exposition purpose.

$$
\begin{aligned}
g(A|U) &= \int_0^\infty \min\left(1, \frac{m}{A}\right) f(m)\, dm \\
&= \int_0^A \frac{m}{A} f(m)\, dm + \int_A^\infty f(m)\, \mathrm{d}m \\
&= \int_0^A \frac{m}{A} f(m)\, dm + 1 - F(A)
\end{aligned}
$$

$$
\begin{aligned}
\frac{dg}{dA} &= \frac{A}{A} f(A) - \int_0^A \frac{m}{A^2} f(m)\, dm - f(A) \\
&= -\int_0^A \frac{m}{A^2} f(m)\, dm
\end{aligned}
$$

B 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 antithetic acceleration. I give brief explanations on their implementation in this Appendix.

B.1 Importance sampling

I employ this technique in order to smooth $g(A_{jt}, m_{jkt}) = \sum_k^{ns} \min\left\{1, \frac{m_{jkt}}{A_{jt}}\right\}$ 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 θ_1 from the first stage, sample again with the following probability:

$$Pr\left(A_{jt} > m_{jkt}|\theta_1\right) = \Upsilon\left(\frac{A_{jt} - \mathbf{D_k}\alpha^m - \beta^m h_{jkt}}{\sigma^m}|\theta_1\right), \tag{B.1}$$

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.

B.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 ζ_{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:

$$\frac{1}{2*ns}\sum_{k}^{ns}\left[\min\left\{1, \frac{m_{jkt}(\zeta_{jkt})}{A_{jt}}\right\} + \min\left\{1, \frac{m_{jkt}(\zeta_{jkt}^{-1})}{A_{jt}}\right\}\right], \tag{B.2}$$

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

B.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: ζ_{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 ϕ_{jt}. Antithetic acceleration is applied when calculating $g(\cdot)$ function, $\frac{1}{2*ns}\sum_k^{ns}\left[\min\left\{1, \frac{m_{jkt}(\zeta_{jkt})}{A_{jt}}\right\} + \min\left\{1, \frac{m_{jkt}(\zeta_{jkt}^{-1})}{A_{jt}}\right\}\right]$.

4. For given values of α^p and β^p, derive residual of inverse demand function, υ_{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 $Pr\left(A_{jt} > m_{jkt}|\theta_1\right)$ which is calculated with the initial estimates.

7. Compute weighting matrix $W = E[Z'\Lambda\Lambda'Z]$.

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).

C Appendix: Description of Variables

References

Anderson, S. P. and Coate, S. (2005) "Market provision of broadcasting: a welfare analysis." *Review of Economic Studies*, 72, pp. 947–972.

Anderson, S. P. and de Palma, A. (2009) "Information congestion." *RAND Journal of Economics*, 40(4), pp. 688–709.

Anderson, S. P. and de Palma, A. (2012) "Competition for attention in the Information (overload) Age." *RAND Journal of Economics*, 43(1), pp. 1–25.

Anderson, S. P. and de Palma, A. (2013) "Shouting to be heard in advertising." *Management Science*, 59(7), pp. 1545–1556.

Anderson, S. P., Foros O., Kind, H. J., and Peitz, M. (2012) "Media market concentration, advertising levels, and ad prices." *International Journal of Industrial Organization*, 30(3), pp. 321–325.

Anderson, S. P. and Gabszewicz, J. J. (2005) "The media and advertising: a tale of two-sided markets." mimeo.

Armstrong, M. (2006) "Competition in two-sided markets." *RAND Journal of Economics*, 37(3), pp. 668–691.

Athey, S., Calvano, E., and Gans, J. S. (2011) "The impact of the Internet on advertising markets for news media." mimeo.

Athey, S. and Gans, J. S. (2010) "The impact of targeting technology on advertising markets and media competition." *American Economic Review*, 100 (20), pp. 608–613.

Bagwell, K. (2007) "The economic analysis of advertising." In M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, Volume 3. Oxford, UK: North Holland.

Berry, S. T. (1994) "Estimating discrete-choice models of product differentiation." *RAND Journal of Economics*, 25(2), pp. 242–262.

Berry, S. T., Levinsohn, J., and Pakes, A. (1995) "Automobile prices in market equilibrium." *Econometrica*, 63, pp. 841–890.

Brown, K. and Alexander, P. J. (2005) "Market structure, viewer welfare, and advertising rates in local broadcast television markets." *Economics Letters*, 86, pp. 331–337.

Butters, G. R. (1977) "Equilibrium Distributions of Sales and Advertising Prices." *Review of Economic Studies*, 44(3), pp. 465–491.

Caillaud, B. and Jullien, B. (2003) "Chicken & egg: competition among intermediation service providers." *RAND Journal of Economics*, 34(2), pp. 309–328.

Chamberlain, G. (1987) "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34, pp. 305–344.

Choi, D. O., Oh, J., Kim, Y, and Hwang, J. (2012) "Competition in the Korean Internet Portal Market: Network Effects, Profits, and Market Efficiency." *Review of Industrial Organization*, 40(1), pp. 51–73.

Clements, M. T. and Ohashi, H. (2005) "Indirect network effects and the product cycle: video games in the U.S., 1994–2002." *Journal of Industrial Economics*, 53(4), pp. 515–542.

Dubé, J.-P. (2005) "Product differentiation and mergers in the carbonated soft drink industry." *Journal of Economics & Management Strategy*, 14(4) pp. 879–904.

George, L. M. and Hogendorn, C. (2012) "Aggregators, Search and the Economics of New Media Institutions." mimeo.

Goeree, M. S. (2008) "Limited Information and Advertising in the U.S. Personal Computer Industry." *Econometrica*, 76(5) pp. 1017–1074.

Grossman, G. M. and Shapiro, C. (1984) "Informative Advertising with Differentiated Products." *Review of Economic Studies*, 51(1) pp. 63–81.

Kaiser, U. and Wright, J. (2006) "Price structure in two-sided markets: evidence from the magazine industry." *International Journal of Industrial Organization*, 24, pp. 1–28.

Nair, H., Chintagunta, P., and Dube, J. P. (2004) "Empirical analysis of indirect network effects in the market for personal digital assistants." *Quantitative Marketing and Economics*, 2, pp. 23–58.

Nevo, A. (2001) "Measuring market power in the ready-to-eat cereal industry." *Econometrica*, 69(2), pp. 307–342.

Petrin, A. (2002) "Quantifying the benefits of new products: the case of the minivan." *Journal of Political Economy*, 110(4), pp. 705–729.

Rochet, J.-C. and Tirole, J. (2003) "Platform competitioin in two-sided markets." *Journal of European Economic Association*, 1(4), pp. 990–1029.

Rysman, M. (2004) "Competition between networks: a study of the market for Yellow Pages." *Review of Economic Studies*, 71, pp. 483–512.

Smith, R. J. and Blundell R. W. (1986) "An exogeneity test for a simultaneous equation Tobit model with an application to labor supply." *Econometrica*, 54(3), pp. 679–685.

Stern, S. (1997) "Simulation-based estimation." *Journal of Economic Literature*, 35(4), pp. 2006–2039.

Van Zandt, T. (2004) “Information overload in a network of targeted communication.” *RAND Journal of Economics*, 35, pp. 542–560.

Wilbur, K. C. (2008) “A two-sided, empirical model of television advertising and viewing markets.” *Marketing Science*, 27(3), pp. 356–378.

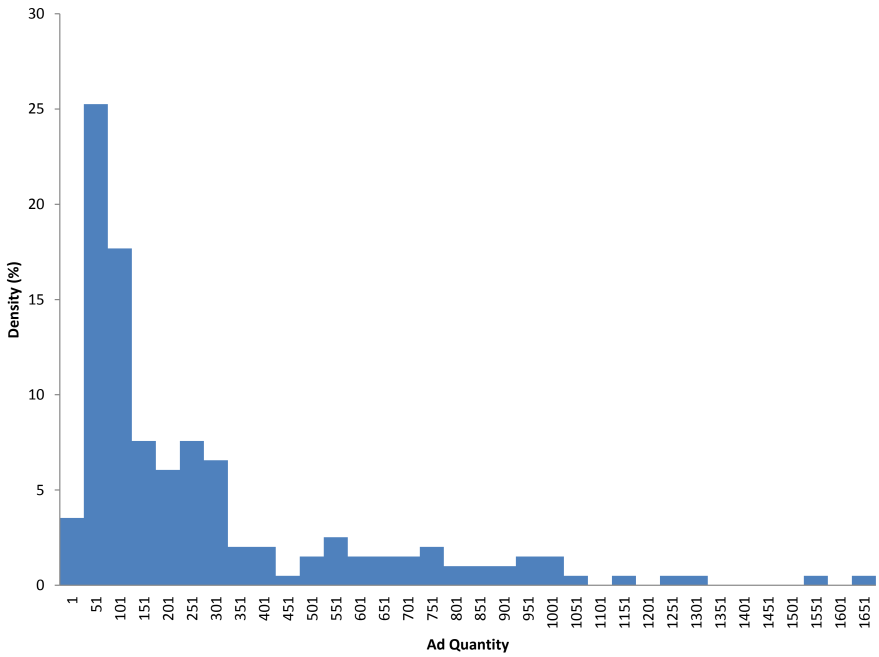

Figure 2: Ad quantity distribution.

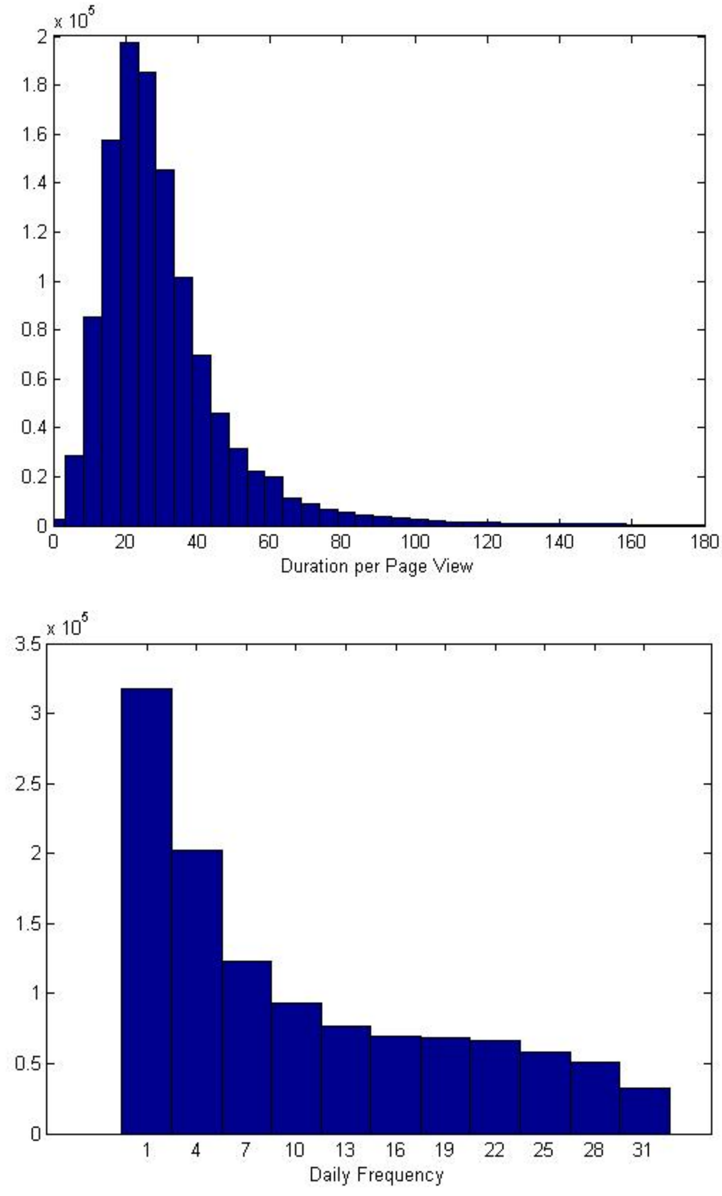

Figure 3: Distribution of Duration per PV and Daily Frequency.

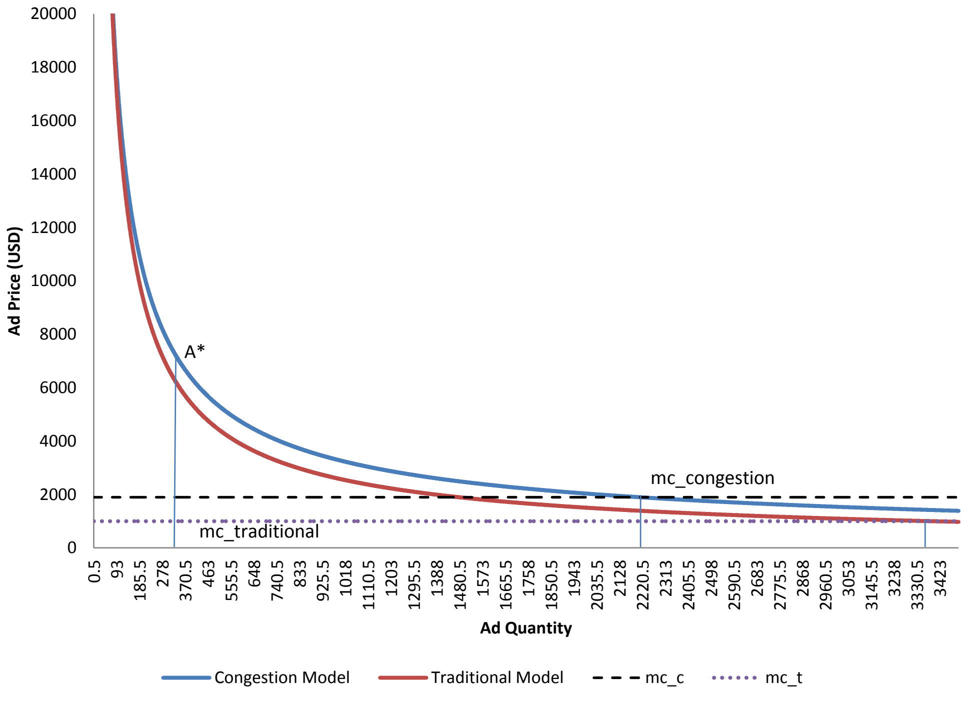

Figure 4: Inverse demand functions for advertising.

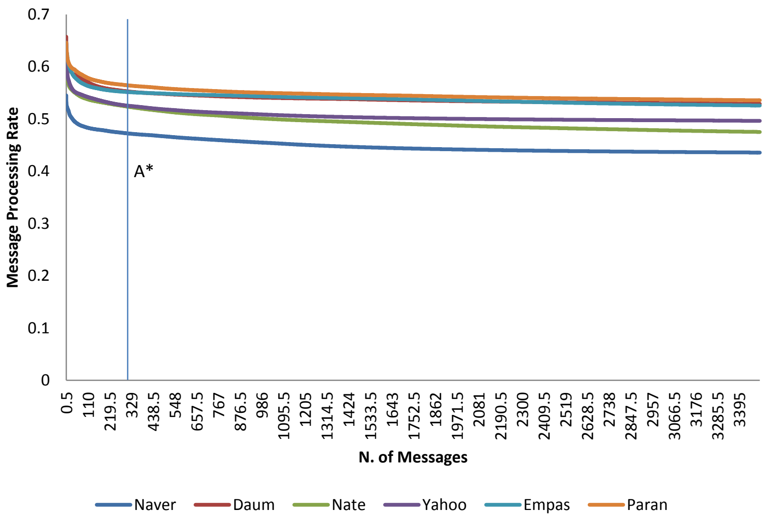

Figure 5: An average attention span of websites according to the ad level.