Asia Pacific Journal of Information Systems
Vol. 31 No. 4 (December 2021), 606-632

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2021.31.4.606

How Content Affects Clicks: A Dynamic Model of Online Content Consumption

Inyoung Chae[a,*], Da Young Kim[b]

[a] *Assistant Professor, Marketing at Sungkyunkwan University, Korea*
[b] *Doctoral Candidate, Emory University's Goizueta Business School, USA*

ABSTRACT

With many consumers being exposed to news via social media platforms, news organizations are challenged to attract visitors and generate revenue during visits to their websites. They therefore need detailed information on how to write articles and headlines to increase visitors' engagement with the content to drive advertising revenues. For those news organizations whose business model depends mainly on advertisements, rather than subscriptions, it is particularly crucial to understand what makes the website attractive to their visitors, what drives users to stay on the website, and what factors affect a user's exit decision. The current research examines individual news consumers' choices to find patterns of increase or decrease in user engagement relative to a variety of topics, as well as to the mood or tone of the content. Using clickstream data from a major news organization, the authors develop a user-level dynamic model of clickstream behavior that takes into account the content of both headlines and stories that visitors read. The authors find that readers appear to exhibit state dependence in the tone of the articles that they read. They also show how the topics expressed in headlines can affect the amount of content readers consume when visiting the news organization to a much larger degree than the topics expressed in the content of the article. Online publishers can make use of such findings to present visitors with content that is likely to maintain and/or increase their engagement and consequently drive advertising revenue.

Keywords: Clickstream analysis, Online content, Topic modeling, Text analysis

Ⅰ. Introduction

Consumers are increasingly turning to online sources as a means of being informed and accessing the news. A recent study conducted by Pew Research Center found respondents aged 18-49 more often get their news online, compared to other sources such as television and radio (Mitchell et al., 2016). Marketing researchers of the news industry have investigated aspects such as media bias (e.g., Xiang

*Corresponding Author. E-mail: inyoung.chae@skku.edu

and Sarvary 2007; Yildirim et al., 2013) and the impact of paywalls (e.g., Chiou and Tucker 2013; Kumar et al., 2012). Recent research has also begun to look at the emergence of echo chambers and the selective consumption of online news (e.g., Del Vicario et al., 2016; Schmidt et al., 2017). However, despite inquiries into aspects of the news industry, those studies mainly focused on the environmental factors of the news industry, rather than which aspects of the articles drive individuals to consume the online news content.

Revenue for news organizations relies on two main sources, subscriptions and advertising. Although print subscriptions continue to decline, news organizations have reported increases in digital subscriptions, notably following the 2016 U.S. Presidential election (Smith, 2017). As an illustration, The New York Times reported an increase of more than 60% in digital-only subscriptions from the end of the second quarter of 2016 to the end of the second quarter of 2017 (The New York Times, 2017). For online content sites in general, advertising revenue is tied directly to the amount of content that individuals consume (e.g., Schweidel and Moe, 2016). With consumers who do not subscribe to news websites, advertising is the only way in which the news organizations can monetize these consumers' visits to their websites. In addition to its subscription revenues, The New York Times also reported that digital advertising revenue in the second quarter of 2017 accounted for $55.2 million, an increase of more than 20% compared to the same quarter from the previous year (The New York Times, 2017). News organizations have a keen interest in understanding those factors that affect the consumption of their content, such as how traffic from social networks drives the revenue associated with both subscriptions and advertising. The amount of time that visitors spend on the website, the number of pages they visit, and the frequency with which they visit the website are metrics often used by news publishers to gauge visitors' engagement (e.g., Cherubini and Nielsen, 2016).

In the current research, we draw on traditional models of clickstream behavior to understand the consumption of news. We make use of a clickstream dataset provided by a major news organization to understand news consumption at the user-level. Our primary interest is in understanding how visitors choose among the set of stories presented to them and identifying any patterns useful for increasing engagement. Various behavioral theories, such as state dependency theory (Heckman, 1991) and variety seeking theory (McAlister, 1982) indicate users' dynamic preference where their prior experience affect subsequent choices. To account for the potential for users' preferences to evolve over the course of a browsing session, we allow for user preferences that drive our choice model to be dynamic. In particular, we assume that visitors' preferences are altered by the news content that they consumed earlier in the same session. We also model the user's decision to continue the browsing session by consuming another story from the set presented to him or to exit the website. Thus, in addition to affecting the next story that the visitor may consume, the news content he consumed earlier in the browsing session may also affect the likelihood that the visitor remains on the site. In this way, we connect visitors' evolving preferences for news content to the number of pages they consume during their session, which is linked directly to the digital advertising revenue the organization generates.

To investigate this issue, we combine text analytic methods (e.g., Blei et al., 2003; Büschken and Allenby, 2016) with clickstream analysis (e.g., Bucklin and Sismeiro, 2003; Moe and Fader, 2004; Montgomery

et al., 2004). We employ latent Dirichlet allocation (LDA), a common topic modeling framework, to characterize the content of headlines and stories and model a user's decision to read another story at the news organization's website and, if so, *which* story, based on the choice set presented to him using a dynamic nested logit model.

Our analysis reveals that users' preferences do in fact evolve over the course of a visit to the news organization's website. We find evidence to suggest that news consumers exhibit a degree of stickiness in the tone of the stories that they read. For example, those who are reading stories about national security are more prone to choose a subsequent story about crime, terrorism or other forms of tragedy. Similarly, we find that a reader of a story about daily activities is more likely to choose an entertainment- or sports-related article, but less likely to choose a story about politics, crime, violence, terrorism, and international tragedies. Choosing stories that have a similar tone to that just read is consistent with research on mood maintenance (e.g., Di Muro and Murray, 2012; Meloy, 2000) and state dependence in online content consumption (e.g., Novak et al., 2003; Schweidel and Moe, 2016).

Given the preference dynamics that news consumers exhibit, the nature of the content that website visitors consume has a bearing on the amount of content they will consume during a browsing session. This has financial consequences for news organizations deriving a portion of their revenue from digital advertising. In the current study, we investigate the effect of changing the prevalence of topics or tone on the duration of sessions.

We provide a review of the related literature in the next section, including research on dynamics in the consumption of digital content and news consumption, followed by a description of the empirical context and the data employed in our research. We then detail the modeling framework, present our findings, and conclude with a discussion of the implications of our research and potential directions for future work.

Ⅱ. Related Literature

To investigate what affects users' online news content consumption, based on the topics of news headlines and stories consumed previously, the current research applies the existing theory of consumers' choice dynamics (i.e., influence of past behavior on future behavior) to the online content publishing sites where each content is represented by (attributed to) a set of topics. In this section, we provide an overview of related studies in consumption dynamics, online clickstream studies, and content analysis.

With the advancement of digital technology and the availability of clickstream datasets that use server logs to identify and record individual user movements, extensive research has investigated consumers' online content consumption. Much of this work largely focuses on the incidence of visits and purchases in commercial sites, and the factors that influence those decisions (e.g., browsing patterns, product types, user-generated content). For example, Moe (2003) investigated shoppers' movements through an online retailer and classified visits as buying, browsing, searching, or knowledge-building. Montgomery et al. (2004) demonstrated how the sequence of pages that a user visits on a website can be informative of the subsequent pages that they will visit. They demonstrated significantly improved accuracy in predicting purchase conversion when leveraging the prior pages visited by the user. Whereas Montgomery et al. examined the dynamics present

within a given browsing session, Moe and Fader (2004) examined how the conversion probability varies across visits to the website by the same user. Bucklin and Sismeiro (2003) showed that website browsing behavior is dynamic, both within a session and across sessions.

Extending these early studies on characteristics of online behavior, the focus of recent research has shifted to investigating the underlying process by which visitors browse websites. For instance, Park and Park (2016) found that visits to a website tend to occur in clusters and that the purchase probability is higher for later visits in a cluster compared to early visits in the cluster. This is in the same spirit of research by Zhang et al. (2015), who demonstrated the importance of considering the clumpiness with which visits may occur at websites. Among the six datasets they considered, they found that clumpiness is the most prevalent at Hulu and YouTube, two online content providers.

Though the focus on incidence of visits and purchases is appropriate for online retailers, it offers limited insights for platforms hosting content, especially for those whose value is primarily generated from the online traffic, such as online portals and social network sites. Though limited, a few studies have recently begun to examine the nature of the content that users are consuming in digital media platforms that deal with digital content such as music and video. For instance, Schweidel and Moe (2016) examined how the breadth and depth of video-viewing sessions at Hulu affect the decision to continue the viewing session and whether to view another video from the same series. They found that, as users view more videos from the same series, they are more prone to remain on the site and view another video from the same series. Datta et al. (2017) examined how the adoption of music streaming platforms affects the volume and breadth of content consumed, finding that users who adopt a music streaming platform not only listen to more music, but also listen to an increased variety of music. Such an understanding of users' content consumption is beneficial for online portals that generate revenue through advertising rather than through transactions.

Despite the long history and significant presence of news in digital media, little research on news consumption has focused on the underlying dynamics, relative to the aforementioned research on other digital content consumption. Much research on the consumption of news has focused on the heterogeneity among news consumers and the competitive nature of the news industry. Mullainathan and Shleifer (2005) demonstrated the importance of heterogeneity among news consumers on the accuracy of the news reported. They assumed that readers want their beliefs to be confirmed and that news organizations can slant the content to do so. Thus, in contexts such as politics where news consumers may have divergent beliefs, news organizations will move toward the extreme. In considering the impact of competition on the news reported, Mullainathan and Shleifer found that more intense competition can result in more extreme slanting. Xiang and Sarvary (2007) further investigated the bias in news by considering the role of conscientious news consumers who seek out more information to discover the truth. Surprisingly, they found that the extent of bias in the news may actually increase with more conscientious consumers, as such consumers may get more information by gathering news from multiple (more biased) outlets rather than from a single source. Simonov and Rao (2017) examined consumer demand for biased news by characterizing different online news outlets based on their ideological positions. In addition to consumers' preference for

news outlets that confirm their ideological positions, they found that the overall quality (e.g., brand power) of the news outlets is a driver of the underlying demand for biased news. Yildirim et al. (2013) examined newspapers' decision to incorporate online editions that include user-generated content. They found that the addition of online editions, while reducing the bias found in the print edition, increases the bias found in online editions due to the presence of user-generated content that is outside the control of the news organization.

In addition to examining the extent of bias that arises from a heterogeneous base of news consumers, researchers have considered the ways in which news is presented to consumers. In examining competition among news programs, Xiang and Soberman (2014) tackled the question of whether organizations should adopt designs that facilitate the delivery of more information in the program. They demonstrated that designs that deliver more information to consumers do not necessarily benefit the news organization and that such designs can provide an advantage over the competition by limiting the benefit that competitors can accrue from improving the design of their news. Roos et al. (2015) investigated whether news websites should provide links to and excerpts from their competitors' websites. While providing links and excerpts associated with another website could be detrimental by making an alternative website more salient to consumers, it can benefit the site that provides the links because consumers may come to rely on the linking website for finding more relevant content.

Recently, several studies have been conducted in other disciplines to understand online news consumption. Consistent with Mullainathan and Shleifer (2005), Del Vicario et al. (2016) found that users diverge in the content that they choose to consume online, resulting in echo chambers that can facilitate the spread of misinformation. By studying two distinct platforms, Bessi et al. (2016) investigated whether this polarization is driven by the content or the algorithms used by online platforms to promote content. They found that the emergence of echo chambers is driven by content, regardless of the platform and its content promotion algorithm. Zollo et al. (2015) examined the dynamics of how misinformation spreads and found a negative trend as the number of comments increases, mirroring the negative trend found in marketing studies of online opinions (e.g., Godes and Silva, 2012; Moe and Schweidel, 2012).

Although there has been substantial research into the dynamics of clickstream behavior in marketing, little work has investigated the dynamics that consumers experience in the consumption of news. Recent studies of the phenomena of misinformation and echo chambers took a macro level view of how stories propagate. The analytical study by Xiang and Soberman (2014) was motivated by the notion that consumers' content consumption occurs in settings similar to that of the current study, with a news provider covering a variety of topics and consumers' interest in certain topics driving their subsequent choices. In their study, though, they assumed that consumers derive their utility from a single, most interesting story, and the choice of news stories is affected by the complexity of the news product, competition, and quantity of information in the news.

By contrast, our intention in this research is to empirically examine the consumption patterns of the *individual* news consumer. To be specific, we directly examine the topics of stories that are reflected in the headlines and actual story content. We allow users' expected utilities to be derived from and evolve based on the content they consume over the course

of a browsing session, which also influences their decision to remain on the website to consume additional content or to end their browsing session. Building on research that has documented the role of consumer heterogeneity in the decisions of news organizations, we incorporate both unobserved heterogeneity across users and temporal variation that consumers may exhibit as they consume news, whether it be to consume more of the same type of news (i.e., positive state dependence) or exhibit an increased preference for a variety of the type of news (i.e., negative state dependence).

The temporal variation over a series of choices investigated in the current study is also closely related to the existing research on the impact of past behavior on consumers' future decisions. Researchers investigating variety seeking have suggested that such consumer behavior may arise due to satiation with product attributes (e.g., McAlister, 1982) or from consumers' desire to balance attributes to maximize utility (e.g., Farquhar and Rao, 1976). Kahn et al. (1986) provide a taxonomy that encompasses both variety-seeking behaviors and reinforcement behaviors. In the context of brand choice, researchers have documented the increased tendency to repurchase the same brand (e.g., Guadagni and Little, 1983), distinguishing between the effects of customer heterogeneity and state dependence (e.g., Fader and Lattin, 1993; Heckman, 1991; Keane, 1997; Roy et al., 1996). Research has also investigated consumption dynamics in the context of services (e.g., Li et al., 2005; Schweidel et al., 2011). In this study, we build upon the existing literature to understand consumers' consumption dynamics in the context of online news consumption.

Ⅲ. Model Development

The goal of this paper is to build a news consumption model that reflects the series of decisions made by consumers when browsing a news website. Hence, we begin by providing an overview of an online user's decision-making processes captured in our modeling approach, and detail the model specification along with the sampling procedure for our hierarchical Bayesian model.

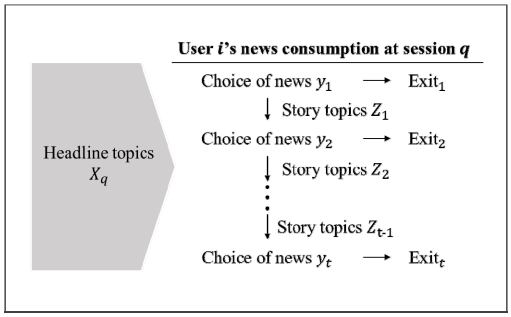
At a news organization's website, consumers often begin their session at the website's homepage or the main page for a section (e.g., U.S. News, World News, Sports, etc.).[1] Both the section main page and the website's homepage present visitors with a list of headlines of recently released articles. From the set of available headlines, visitors choose which article they want to read. After reading a given article, they decide whether to remain on the website and which story to read next, or they choose to end their session at the website. This decision is based on the set of articles with which they are presented at that time, which includes both recent articles and older articles that are related to the content of the story they have just read. Since the related older articles vary over time, across individuals and based on the content just viewed, we treat them as the outside option in our choice model and allow for heterogeneity across individuals in their tendency to choose these stories compared to more recently published stories.

In modeling a visitor's choice among news articles, we recognize that the decision of what to read next may be affected by what the visitor has read earlier

1) Users may also initiate their session at the website by directly visiting a particular story, which would occur if they are referred from a link in an email or on a social media platform. In such instances, we model their subsequent navigation decisions.

in the same session. For example, if a visitor has just read an article on a particular topic, he may be more interested in reading another article on the topic, exhibiting a form of positive state dependence in terms of the content he consumes. Alternatively, he may have satiated on that topic and prefer to read an article on a different topic, akin to exhibiting a preference for variety. To capture such dynamics in visitors' preferences, we allow for the weights associated with the topics present in the headlines of articles to evolve based on the content of the article that a visitor has just read.

In addition to affecting a visitor's choice among articles, the headlines presented may also affect a visitor's decision to remain on the website to read another article or to leave the website. To model a visitor's decision to leave the website, we adopt a measure of the overall attractiveness of the set of articles with which the visitor is presented. Akin to using the attractiveness of items within an individual category to model the category purchase decision (e.g., Bell and Lattin, 1998), we operationalize the attractiveness of the news website based on the attractiveness of the set of headlines presented to a given visitor at a particular point in time. The measure is derived from the expected utility of a nested logit model of article choice, conditional on the previously viewed articles. As the articles that



<Figure 1> Overview of the Proposed Model

a visitor has previously read on the website may affect the attractiveness of each article from which the visitor may choose, previously consumed content may not only affect the next story that a visitor decides to read, but it may ultimately affect the visitor's decision to remain on the website. <Figure 1> demonstrates the main news consumers' decisions modeled.

Choice of Headlines. We define y_{iqt} as a choice of news article made by individual i at session q and incidence t, and it takes a value of $r \in \{R_{iqt}\}$, which is a choice set of articles possibly selected at a given time. The probability to select the article r is specified by:

$$\Pr(y_{iqt} = r) = \frac{\exp(U_{iqtr})}{\exp(U_{io}) + \sum_{j \in \{R_{iqt}\}} \exp(U_{iqtj})} \quad (1)$$

where U_{iqtr} is the deterministic utility for each option of article. We further specify this utility as a function of topics of headlines that users can observe from the homepage or front pages of subsections (e.g., Politics, Business, Entertainment, etc.):

$$U_{iqtr} = X_{iqtr}\beta_{iqt} \text{ for } r \in \{R_{iqt}\} \quad (2)$$

Here, X_{iqtr} is P-covariate vector that includes the distribution of topics in article r's headline. To obtain this topic distribution, as we discuss in more detail in the next section, we preprocessed the entire set of headlines using the LDA model, selected the number of topics manifested (P + 1), and calculated the first-difference values based on the (P + 1)-th topic weight. In other words, the LDA estimates provide a vector of topic probabilities $([x_1 \ \cdots \ x_p x_{P+1}])$ that characterize the headline of each article. To avoid multi-collinearity, the topic distribution is reconstructed as first-difference values (i.e., $[x_1 - x_{P+1} \ \cdots \ x_p - x_{P+1}]$). β_{iqt} is the corresponding coefficient vector that weights each topic

to drive an article viewing decision.

The probability of choosing the outside option is addressed by the individual-specific baseline U_{i0}. Rather than assuming a baseline utility of 0 for all individuals, we specify a random individual baseline as $U_{i0} \sim N(0, \sigma^2)$ and allow it to persist across sessions and incidents within a session. This will capture the tendency of each individual to consume older stories compared to the newly updated articles. Thus, the probability of choosing an article that does not appear in the front page is given by:

$$\Pr(y_{iqt} = r) = \frac{\exp(U_{i0})}{\exp(U_{i0}) + \sum_{j \in \{R_{iqt}\}} \exp(U_{iqtj})} \quad (3)$$

Dynamic effect of previous article. The content of articles that visitors viewed previously can affect their following viewing decisions. To address this, we retrieved the topics of each article's content by implementing LDA with S+1 topics.[2] We derived first-differenced values (Z_{iqtr}) and allow these values to affect preferences for the topics contained in article headlines:

$$\beta_{iq(t+1)} = \gamma_0 + Z_{iqt}\gamma \quad (4)$$

Exit decision. Lastly, a visitor's decision of whether to terminate their browsing session or to remain on the website and view more articles is captured by our measure of overall website attractiveness. We assume that website attractiveness is given by the expected utility of a nested logit model (e.g., Bell and Lattin, 1998):

$$WA_{iqt} = \ln[\exp(U_{i0}) + \sum_{j \in R_{iqt}} \exp(U_{iqtj})] \quad (5)$$

The probability with which visitor i chooses to exit the website at session q and incidence t is:

$$\Pr(exit) = \frac{1}{1+\exp(V_{iqt})} \quad (6)$$

where:

$$V_{iqt} = \delta_0 + \delta_1 WA_{iqt} \quad (7)$$

The overall likelihood of the model can be specified as:

$$\Pr(y_{iqt} = r) = \left[\frac{1}{1+\exp(V_{iqt})}\right]^{I(r=0)} \left[\frac{\exp(V_{iqt})}{1+\exp(V_{iqt})} \times \left\{\frac{\exp(U_{iqtr})}{\exp(U_{i0}) + \sum_{j \in \{R_{iqt}\}} \exp(U_{iqtj})}\right\}^{I(r \in \{R_{iqt}\})} \left\{\frac{\exp(U_{i0})}{\exp(U_{i0}) + \sum_{j \in \{R_{iqt}\}} \exp(U_{iqtj})}\right\}^{I(r \notin \{R_{iqt}\})}\right]^{I(r>0)} \quad (8)$$

where $r = 0$ if individual i chooses to exit the website, and $r \in \{R_{iqt}\}$ if he remains on the site and chooses one of the articles with which he is presented on incidence t of session q.

For estimation convenience, we incorporate the dynamic process of β_{iqtr} into the model as follows:

$$U_{iqt} = \boldsymbol{X}_{iqt}\Gamma\boldsymbol{Z}_{iq(t-1)} \quad (9)$$

where X_{iqt} is a $(R_{iqt} \times P)$ matrix of the topic distribution of the headlines in visitor i's consideration set for incidence t of session q, $\boldsymbol{Z}_{iq(t-1)}$ is

2)These S+1 topics derived from news stories are distinct from the P+1 topics identified from the headlines of the articles, allowing the sets of topics characterizing the headlines to differ from those of the article content. Doing so provides increased flexibility to test empirically which topics of headlines and stories are indeed critical in the users' news consumption decisions.

a (S+1) vector of $[1\ z_{iq(t-1)}]$, and Γ is $(S+1) \times P$ coefficient matrix that captures the dynamic evolution between topic preferences and story consumption.

The model is estimated using a fully Bayesian approach. To ensure a well-identified model with proper posterior determined almost entirely by the data, we assign non-informative diffuse distributions. For the coefficient parameters, we adopt normal priors: $\Gamma_p \sim MVN(\mu_\gamma = 0, \Pi_\gamma = 100I_{s+1})$ and $\delta \sim MVN(\mu_\delta = 0, \Pi_\delta = 100I_2)$. For variance parameter σ^2, we adopt inverse-gamma prior: $\sigma \sim IG(n_0 = 0.1, D_0 = 0.1)$. Under prior independence, the corresponding joint posterior is given by[3]:

$$\begin{aligned}&\pi(\Gamma, \delta, \sigma | y, \boldsymbol{X}_{iqt}, \boldsymbol{Z}_{iq(t-1)}) \propto \\ &\prod_{i=1}^{N}\Big[\prod_{q=1}^{Q_i}\prod_{t=1}^{T_{iq}} f(y_{it} = exit|\ \delta)^{I(r=0)} \times \\ &\{f(y_{it} \neq exit|\ \delta))f(y_{it} = r|\Gamma, \delta, U_{i0})\}^{I(r>0)}\pi(\Gamma)\pi(\delta)\Big] \\ &f(U_{i0}|\sigma^2)\pi(\sigma^2)\end{aligned} \quad (10)$$

Ⅳ. Empirical Analysis

4.1. Data Description

The data for the current study comes from a global news organization that provides news content covering politics, business, crime, sports, entertainment, and more to an audience of more than 100 million unique visitors each month. Unlike competing news publishers that impose paywalls or restrictions to access their digital content, the organization's digital revenue is through the delivery of advertisements including banners, videos, and interactives. Therefore, it is crucial for management to understand the behavior of their online visitors in terms of how they consume content and what drives these users to stay on the website.

Our dataset consists of a series of desktop (non-mobile) users' clickstreams from April 1st to July 31st, 2015. It includes unique IDs for each user and session, the URLs of the web pages and their news categories as assigned by the news organization, and corresponding time stamps of the visits. To garner managerially relevant yet estimable information, we selected a cohort of users who had visited the website at least once in April 2015. In this way, the sampled visitors are likely to be regular users who had constantly consumed news articles depending on the articles' characteristics, rather than those who are new to the website or those who have not visited for a while. Among this cohort, we removed users who visited 1) pages other than the homepage and section main pages for which the publication dates were not identifiable for choice set formation, 2) pages where the html files for the headline and story texts were not available, and 3) non-English pages (to keep the language consistent when extracting text for the topics). Lastly, we excluded bouncers (e.g., users who only scan through headlines in the front-page and exit the website afterwards or those who accidently visit the site without further engagement) due to their lack of individual-level choices.

After screening the cohort based on the above criteria, we randomly sampled 5,112 unique users and their recurring visits to the news website for the last three months of the observation period (May 1st to July 31st). The sample dataset includes 84,091 sessions and 164,077 page-views in total. This converts to an average of 32.1 sessions per visitor (s.d. 69.6) and 2.0 page-views per session (s.d. 1.7) over

3) We provide the details of the MCMC algorithm used for the estimation in the Appendix.

the three months (<Table 1>). Each session is assigned by the data provider as a continuous chain of page-views that has less than 30 minutes of break between page views (e.g., Bucklin and Sismeiro, 2003).

For our model estimation, we extended this clickstream dataset and constructed two types of datasets: user-focused news consumption data and article-focused topic data. The first dataset comprises the sampled users' page-level clickstream dataset with their choice set information at the point of click. Although we collected enough information about the users' browsing decisions, we needed to construct the choice set, because the raw clickstream does not include the exact image of the homepage where headlines were displayed at the time of each user's clicking. Our exploratory analysis of the data revealed that more than half of the articles displayed on the homepage were published on the same date. Our discussions with the organization revealed that those stories appearing on the homepage after their publication date are related to special circumstances (e.g., holidays, minor news sections). Thus, we constructed the consideration set for a given incidence as the set of article headlines that were published on the same day. To be specific, if a user started browsing from the homepage, the corresponding consideration set includes every article headline that was published on the same day (87% of all consideration sets), whereas if a user started browsing from a section main page (13%), we constructed the set of article headlines that were published under that specific news section on the same day.

Furthermore, we updated the choice sets for each page view by excluding previously visited pages within the same session from the choice sets. Unlike shopping clickstream data where users visit the same product page repeatedly before making a purchase decision, the nature of the news media makes it unlikely for customers to return to an article they have already read. Lastly, pages that did not contain publication dates (e.g., program information pages, help pages, administrative pages; 8.07% of the unique web pages), or pages that were published before the visit date, were omitted from the choice set and considered outside options. The summary of the final choice set is presented in <Table 2>.

The second dataset consists of characteristics of the articles that have been read by the users or displayed in their choice set at each browsing occasion. The most convenient and commonly used way to

<Table 1> Descriptive Statistics for Users' Clickstream Data

Variable	Mean	Std dev	Min	Max
Number of sessions per user	32.10	69.55	1	2,201
Number of page views per session	1.95	1.69	1	45

<Table 2. Summary of Choice Sets

Section	Mean	*SD*	Section	Mean	*SD*	Section	Mean	*SD*
Homepage	169.16	61.24	Opinion	6.24	3.49	Justice	3.40	3.84
US	46.14	18.31	Sports	6.64	4.61	User reports	3.00	2.94
World	32.62	15.07	Health	4.07	3.51	Tech	1.32	1.76
Politics	25.66	9.99	Travel	4.21	2.85	Business	1.36	1.45
TV	20.57	11.40	Style	2.17	1.70	Videos	1.67	2.00
Entertainment	7.42	4.33	Living	3.74	3.19	Other	2.78	2.97

characterize articles would be to use the topics assigned by the publisher (e.g., Song et al., 2016) or classify based on content formats (e.g., UGC vs. PGC [Chae et al., 2017], articles vs. videos, published content vs. external links [Roos et al., 2015]). For example, most news sites categorize articles by topic (e.g., politics, world news, business, culture) or by format (e.g., forums, blogs, photo albums, videos). Such structured information may not capture the actual consumers' choice mechanism when the choice is made within the broadly defined categories. Consumers may not in fact choose an article to read solely based on those categories; rather, they may skim a list of articles displayed and choose the ones to read based on a specific topic in which they are interested. Furthermore, from a model estimation point of view, using such categories does not differentiate articles that belong to the same category, and in turn, results in loss of information. Therefore, instead of the structured categories, we use text analysis to characterize each article and corresponding story to estimate our news consumption model.

Our text analysis follows the next four steps. First, using the web URLs in the clickstream dataset, we crawled corresponding articles and processed them as headline and story content. This results in 49,756 headlines (the average number of words per headline: 3.6) and 45,757 stories (the average number of words per story: 180.0)[4]. Next, we preprocessed each headline and story dataset by eliminating overly specific components such as numbers and punctuation, stripping out white space, removing stop words (using the SMART information retrieval system [Salton 1971]), removing sparse terms (words that are observed less than 0.1% in the entire word set are excluded), and stemming words. As a result, we have 1,192 and 9,884 unique words in the headline and story datasets, respectively.

Third, we use the LDA model to define each article's headline and story based on the topics of the content (Blei et al., 2003). By running a LDA model on each dataset, we can estimate the underlying set of latent topics, each of which is represented by a distribution of observed words (indexed by β in Blei et al. (2003)'s model); an underlying topic distribution for each article (indexed by θ); and a latent topic used in each word in an article (indexed by z). Conceptually in our research setting, we assume that there are topic distributions captured from the entire set of news articles, with the news contributor having certain intent (represented as a mixture of topics) for each article and using specific sets of words to express the topics in the article. A goal of our model is to understand users' news consumption choices among articles, each of which is characterized by its latent topics, discovered from the LDA model, that represent and distinguish one article from another. Specifically, we use the underlying topic distributions (θ), which characterize each article's topic and words, for the content characteristics variable.

In addition, considering that each topic is defined by a mixture distribution of words that are observed from the entire dataset in the LDA model, we separate the headline and story dataset so that each headline and story can be represented only by relevant words in the respective headlines and stories. This process is necessary due to the difference in the nature of the words used; words used for headlines are more general and abstract compared to stories, whereas words used within stories are likely to be more specific and varied. Consequently, the LDA model allows us to characterize each article's headline and story

4) The difference in the number of headlines and stories is due to the nature of the pages that do not involve stories (e.g., galleries, videos, infographics, etc).

as an underlying set of topics. These topic distributions are used in the news consumption model as a set of attributes, which is equivalent to product attributes in a product choice model (e.g., Fader and Hardie, 1996; Gilbride and Allenby, 2004; Guadagni and Little, 1983; Kim et al., 2007; Lattin, 1987).

Since the values of topic probabilities are summed to one for each article, we define the smallest topics in terms of the proportion captured in the respective datasets (headline set and story set) as the "baseline topic" and use the first difference as covariates. Similar to selecting variables in a way to maximize model fit in choice models (e.g., Gilbride et al., 2006), we adopt the information criteria and select variables (i.e., deciding the number of topics) that maximize the criteria generated from our news consumption model. Alternatively, we can also consider the model with the number of topics that captures the articles' textual information the best (i.e., using perplexity measure) based on LDA model performance. However, this method is more text-oriented rather than consumer-oriented. As our goal in this research is to understand visitors' news consumption behavior, we select the model that performs the best at capturing consumers' news choices. The next section demonstrates the outcome of the fit tests using different model specifications and details the final number of topics selected from the best fitting model.

4.2. Model Selection

To find the best performing model, we examine the model fit by calculating the log-marginal density (LMD; Newton and Raftery, 1994) and the deviance information criterion (DIC) across various combinations of the number of headline and story topics. Both measures capture goodness fit of models to observed outcomes. The LMD is the standard Bayesian model metric, whereas DIC takes into account the impact of increasing number of parameters by penalizing the fit measure. We extensively explore the fit statistics for candidate models that jointly vary over the number of headline topics (5 – 18 topics) and the number of story topics (4 – 14 topics); we found the best performance in terms of LMD and DIC with the 14 headline topics – 6 story topics pair compared to other candidates. Thus, we proceed with the following empirical analysis based on these 14 headline topics and 6 story topics. The results are demonstrated in <Table 3>.

Based on the best fitting model, called the "14 headline 6 story model," we listed the prevalent words that have higher occurrence probabilities for each latent topic (<Table 4> and <Table 5>) and labeled the topics based on those words (listed in the first rows of each table). The second rows of the tables show how much of those topics are captured across articles. These values show that except for the topic "*Website features,*" the headline topics across documents are comparable to the average proportions lying between 5.6% and 9.1%, whereas the proportions of story topics are more unbalanced, with the topic of "*Crime*" having the largest share (31.8%) on average, followed by "*Daily Life* (21.8%)."

From topics extracted from the article headlines (<Table 4>), we reveal that some topics are specific to the events that occurred during the observation period (e.g., "*Social issues/ Court rulings,*" "*Politics/ Election,*" "*Entertainment/ Sports,*" "*Natural disasters,*" "*Police brutality*"), while others deal with ordinary news but are specific enough to capture detailed information (e.g., "*Gun violence,*" "*Crime,*" "*World news,*" "*ISIS/ Terrorism,*" "*Health/ Home*"). For instance, the topic "*Entertainment/ Sports,*" which includes keywords on celebrities, TV shows, and sports

<Table 3> Model Fit Comparison

		Story 4	Story 5	Story 6	Story 7	Story 8
Headline10	LMD	-219372.7	-219349.5	-219401.2	-219363.5	-219427.6
	DIC	717785.3	717762.0	717699.8	717677.6	717651.9
Headline13	LMD	-218697.9	-218722.1	-218661.7	-218674.2	-218648
	DIC	714935.9	714928.8	714905.8	714858.5	714951.7
Headline14	LMD	-218917.5	-218982.9	**-216711.7**	-219000.5	-219012.3
	DIC	715822.1	716163.4	**714864.7**	716048.0	716146.7
Headline15	LMD	-218891.8	-218881.8	-218874.1	-218861.8	-218869.5
	DIC	715808.7	715781.6	715716.8	715683.1	715744.4
Headline18	LMD	-218730.9	-218728.4	-218734.5	-218754.1	-218723.4
	DIC	715136.8	715171.1	715181.7	715095.8	715151.1

<Table 4> Description of Headline Topics and Primary Words

Topic	Social issues / Court rulings	Family tragedy	Gun violence	Politics / Election	Entertainment / Sports	Crime	World news
	9.1%	8.7%	8.1%	7.9%	7.7%	7.6%	7.5%
1	marriage	man	video	Clinton	star	prison	world
2	law	baby	police	Hillary	show	charge	pope
3	court	kill	shoot	Iran	top	murder	President
4	gay	mom	kill	Obama	David	death	China
5	same sex	die	shot	deal	final	escape	meet
6	flag	found	suspect	Bush	win	case	climate
7	immigration	daughter	caught	President	secret	accuse	Russia
8	confederate	son	camera	poll	actor	rape	state
9	supreme	family	charge	campaign	Jenner	guilty	Cuba
10	religion	save	release	GOP	Letterman	drug	Francis
Topic	**ISIS / Terrorism**	**Natural disasters**	**Police brutality**	**Transportation accidents**	**North Korea / Mass murder**	**Health / Home**	**Website features**
	7.3%	7.2%	7.1%	6.7%	5.8%	5.6%	2.9%
1	ISIS	Nepal	Baltimore	plane	North	kid	[News Organization]
2	attack	earthquake	Gray	crash	Korea	study	[News Organization].com
3	terror	flood	protest	flight	Boston	life	news
4	Syria	California	Freddie	train	bomb	cancer	profile
5	Iraq	found	arrest	pilot	victim	health	transcript
6	fight	Texas	report	airline	attack	disease	correspond
7	military	tornado	Ferguson	Amtrak	Tsarnaev	teen	report
8	force	hit	case	air	marathon	brain	anchor
9	threat	coast	riot	jet	Carolina	food	digital
10	war	rescue	investigate	passenger	trial	parent	produce

results, also contains the words "David" and "Letterman" following his retirement from the *Late Show* on May 20th, 2015 and "(Caitlyn) Jenner" after her identification of herself as transgender in April

2015. Likewise, reflecting the news stream of the upcoming presidential election, the topic "*Politics/ Election*" consists of the names of then-president "Obama" as well as the names of the presidential primary candidates, such as "Hillary," "Clinton," or "Bush," and election keywords such as "poll" and "campaign." Words that describe the topic "*Social issues/ Court rulings*" are also associated with the social issues in the United States at the time, including the Supreme Court ruling on same-sex marriage and opposition from some religious organizations, and the tensions over the confederate flag and acts of white supremacy.

Some topics provide a more detailed account of the news material than may have been assigned to a single section under the news website's categorization. News on accidents and disasters fall under the topics "*Natural disasters*" and "*Transportation accidents,*" where words such as "earthquake," "flood," and "tornado" relate to natural disasters, whereas "plane," "train," and "crash" refer to accidents with transportation vehicles. We also observe everyday headlines that would have been categorized under a typical news section that will contain subtle distinctions in terms of the topics dealt with in the articles. For instance, while the topics "*Gun violence,*" "*Crime,*" "*Police brutality,*" and "*North Korea/ Mass murder*" fall under the umbrella of criminal events, the topic "*Police brutality*" specifically depicts headlines related to reactions toward police brutality, with the words "Baltimore," "Ferguson," "riot," "protest," and "Freddie" that are specific to the nation-wide protests in reaction to the deaths of Michael Brown and Freddie Gray. Headlines related to gun violence are covered by "*Gun violence,*" containing the words "police," "officer," "shoot," and "suspect," as well as the words "caught," "video," and "camera" which portray the website's use of video clips sourced from the witnesses of such occurrences. "*North Korea/ Mass murder*" largely revolved around two issues at the time: the Boston marathon bombing ("Boston," "bomb," "Tsarnaev") and North Korea's announcements on nuclear missile developments ("north," "Korea," "attack," "Kim"). The headlines of general crime articles can be described with words such as "prison," "drug," "rape," and "accuse" under the topic "*Crime.*"

<Table 5> displays the keywords for the six topics that compose the story text of the news articles. The topic "*Crime*" involves words related to crime, including "police," "investigation," and "arrest." The topic "*Daily life*" consists of words about life, relationships, and emotions, such as "life," "family," "friend," "love," and "play." The story texts also include a share of the words related to politics, with the words "Washington," "President," "Senate," "House," "Republican," and "Democrat." Government decisions and actions on national security are described under the topic "*National security*" with "military," "attack," and "war," whereas general words on international issues are under the topic "*International,*" such as "world," "travel," and "people."

Interestingly, in contrast to an expectation that the number of topics would likely be greater for stories that have more content than a headline, which is often a single line of text for each article, our model finds that a smaller number of story topics matter in the news consumption model. We suspect that this is because our proposed model only captures the topics that indeed influence visitors' news consumption patterns, rather than capturing all topics that represent the content of articles. In other words, although the volume of content is larger in stories compared to that in headlines, visitors' news consumption behaviors might be affected by a smaller set of story topics compared to headline topics that directly affect clicks to the next articles. Keeping

<Table 5> Description of Story Topics and Primary Words

Topic	Crime	Daily life	International	Politics	National security	Health/Science
	31.8%	21.8%	13.6%	13.0%	11.2%	8.6%
1	report	year	world	President	govern	study
2	police	time	city	state	state	problem
3	investigate	life	people	Republican	unit	health
4	charge	love	area	Washington	military	find
5	kill	make	time	politics	force	research
6	death	family	travel	house	nation	increase
7	case	play	place	Democrat	secure	live
8	found	work	local	issue	attack	state
9	arrest	friend	nation	campaign	leader	change
10	attorney	story	land	senate	war	number

these implications in mind, we will elaborate the findings of our news consumption model in the next section.

4.3. Model Results

As mentioned previously, for identification, we use the least prevalent topics for each model as a baseline topic ("*Website features*" for the headline topics and "*Health/Science*" for the story topics) and use the first difference of the rest of the topics and the baseline topic (i.e., $Covariate_j = \%\ of\ topic_j - \%\ of\ baseline\ topic$ as the main covariates of the model. In this section, we will demonstrate which topics affect consumers' news consumption decisions based on headline topics (the second column of <Table 6>), how topics in the stories chosen to read affect the next choice of news consumption (3-7 columns in <Table 6>), and how the overall attractiveness of news articles presented on that day affects consumers' exit decisions. Finally, we use a simulation study to examine the overall effect of each topic in terms of an impact to the length of session (i.e., the volume of news consumed in a session), which ultimately matters to the news publisher.

First, our proposed model yields estimates for the impact of the headline topics on the news consumption decisions. The second column in <Table 6> presents the posterior mean of parameter γ_0 for each headline topic and its 95% confidence interval. The results show that 12 out of 13 topics (excluding the baseline topic "*Website features*") capture significant effects (either positive or negative), whereas only the topic "*Entertainment/ Sports*" captures an insignificant effect. For example, if a news articles' headline consists of "*Crime,*" a news reader is three times more likely to choose the article compared to an exit option $(\exp(\hat{\gamma}_0^{Crime}) = 3.06)$, where as he is 41% less likely to choose the article if it is composed of topics related to "*World news*" $(\exp(\hat{\gamma}_0^{World\ News}) = -0.59)$. Overall, if the news is associated with violent, upsetting, and stimulating topics, such as "*Crime,*" "*Gun violence,*" "*Natural disasters,*" "*Transportation accidents,*" "*Police brutality,*" "*Mass murder,*" "*Terrorism,*" and "*Family tragedy,*"

<Table 6> Preferences for Headline Topcs and Story Topics' Dynamic Effect

(to) Headline Topics	Baseline	(from) Story Topics				
		Crime	Daily life	International	Politics	National security
Social issues / Court rulings	-0.11** (-0.17, -0.06)	-0.17 (-0.40, 0.08)	-0.12 (-0.40, 0.16)	-0.34 (-0.69, 0.06)	0.59** (0.29, 0.89)	-0.18 (-0.57, 0.16)
Family tragedy	0.20** (0.15, 0.26)	0.13 (-0.08, 0.31)	0.08 (-0.17, 0.31)	0.06 (-0.22, 0.38)	-0.10 (-0.41, 0.23)	0.39** (0.10, 0.67)
Gun violence	0.72** (0.67, 0.77)	-0.10 (-0.28, 0.08)	-0.29** (-0.50, -0.08)	0.05 (-0.23, 0.34)	0.07 (-0.20, 0.36)	0.08 (-0.20, 0.35)
Politics / Election	-0.41** (-0.49, -0.34)	0.07 (-0.19, 0.33)	-0.33** (-0.67, -0.03)	-0.43 (-0.89, 0.02)	0.88** (0.56, 1.18)	0.10 (-0.34, 0.53)
Entertainment / Sports	0.06 (-0.01, 0.13)	0.00 (-0.25, 0.22)	0.40** (0.17, 0.66)	-0.65** (-1.05, -0.23)	0.16 (-0.19, 0.47)	0.11 (-0.31, 0.45)
Crime	1.12** (1.08, 1.17)	-0.11 (-0.28, 0.07)	-0.46** (-0.66, -0.27)	-0.21 (-0.47, 0.09)	-0.01 (-0.25, 0.24)	0.41** (0.16, 0.67)
World news	-0.53** (-0.61, -0.45)	0.04 (-0.27, 0.34)	0.16 (-0.12, 0.55)	0.05 (-0.49, 0.54)	0.09 (-0.37, 0.55)	0.36 (-0.07, 0.75)
ISIS / Terrorism	0.26** (0.20, 0.32)	-0.10 (-0.38, 0.15)	-0.30 (-0.58, 0.01)	0.36** (0.03, 0.69)	-0.06 (-0.43, 0.35)	0.66** (0.38, 0.93)
Natural disasters	0.52** (0.47, 0.58)	0.01 (-0.21, 0.22)	-0.44** (-0.68, -0.18)	0.19 (-0.14, 0.49)	0.23 (-0.12, 0.52)	-0.02 (-0.31, 0.27)
Police brutality	0.39** (0.33, 0.46)	-0.01 (-0.25, 0.21)	-0.35** (-0.65, -0.04)	-0.48** (-0.92, -0.10)	0.23 (-0.14, 0.53)	0.39** (0.06, 0.71)
Transportation accidents	0.50** (0.45, 0.57)	-0.17 (-0.38, 0.05)	-0.41** (-0.70, -0.13)	0.09 (-0.19, 0.41)	-0.33 (-0.69, 0.03)	0.41** (0.11, 0.69)
North Korea / Mass murder	0.38** (0.31, 0.44)	0.03 (-0.22, 0.29)	-0.37** (-0.67, -0.07)	0.08 (-0.29, 0.44)	-0.32 (-0.74, 0.09)	0.30 (-0.06, 0.66)
Health / Home	-0.45** (-0.53, -0.37)	0.14 (-0.25, 0.50)	0.30 (-0.09, 0.69)	-0.52** (-1.05, -0.01)	0.37 (-0.05, 0.79)	-0.57** (-1.21, 0.00)

Note. The 95% credible intervals are reported in parentheses. **0 is not contained in the 95% credible interval.

it is more likely to be selected to read ($\hat{\gamma}_0 > 0$ for such topics). On the other hand, if headlines of the articles cover mundane and everyday life-related topics, such as "*Social Issues,*" "*Politics,*" "*Health,*" and "*World News,*" those articles seem to lose news viewers' attention ($\hat{\gamma}_0 < 0$ for such topics). This finding of consumers' preference for stimulating and violent content over mundane content is consistent with studies on the impact of arousing content on virality by Berger and Milkman (2012). It is plausible that the graphic descriptions of real-time tragic events on the news media would stir anger and anxiety in the minds of the visitors, in turn drawing more attention from the users compared to mundane topics that cover day-to-day updates on domestic and international issues or informational contents on health and family. The general appeal of tragic contents is also in line with research in the media and commu-

nication disciplines (Ahn et al., 2012; Goldenberg et al., 1999; Mares and Cantor, 1992). The reasons behind this pursuit of tragedy is found to be related to one's empathy towards the victims, to building one's coping mechanism towards tragic events, or to one's downward comparisons with subjects who are worse off than themselves.

One of the key features of our modeling framework is a dynamic component that captures the way in which content preferences for the choice of which headline to read next are updated based on the story content previously viewed. Columns 3 – 7 in <Table 6> report the results of these dynamic components of the model. Unlike the baseline preferences where 92% of coefficients were significant, the results of dynamic components show a strong alignment between headline topics and story topics, with 31% of coefficients for story topics being significant. More interestingly, the results can largely be interpreted with topical and tonal relevance between the headline and story and with the domestic/international scope of the topics. For example, after reading stories that contain the topic "*Politics,*" readers are more likely to click on headlines of "*Politics / Election*"(60% more) and/or "*Social issues / Court rulings*" topics (62% more) compared to the probability of selecting an outside option. This change is especially dramatic considering that the baseline preference of headlines related to such topics is negative, $\hat{\gamma}_0^{Politics} = -0.41$ and $\hat{\gamma}_0^{Social} = -0.11$, respectively, but that the probability of consuming those topics becomes positive after reading the "*Politics*" topics. We also find that users who have finished reading an article on international issues are more likely to remain with their international scope with headlines related to "*ISIS / Terrorism,*" rather than focusing on the domestic news on "*Police brutality,*" "*Health/Home,*" and "*Entertainment / Sports.*" The propensity to select the "*ISIS*"-related headlines is accentuated by 43%, whereas the other three topics-related headlines are decreased by 48%, 41%, and 48%, respectively[5].

We also observe users' tendency to keep the mood consistent across their reading selections. Users who read stories that discuss topics on national security are more likely to choose article headlines that contain grave and serious topics – e.g., "*ISIS/Terrorism*" (93% increase), "*Crime*" (51% increase), "*Transportation accidents*" (51% increase), "*Family tragedy*" (48% increase), "*Police brutality*" (48% increase) – and move away from educational or lighthearted headlines, such as "*Health/Home*" (43% decrease). By contrast, those who went through a story on the "*Daily life*" topic stick to light and entertaining headline topics and avoid headlines related to dreadful and vicious topics. Such readers are specifically prone to choose articles with "*Entertainment/ Sports*" headlines (49% more), while staying away from articles with headlines nuanced with crime and misfortune (e.g., "*Crime,*" "*Gun violence,*" "*Police brutality,*" "*North Korea/Mass murder,*" "*National security,*" "*Transportation accidents,*" "*Natural disasters,*" "*Politics/Election*"). On average, headlines with such topics are 31% less likely to be selected compared to the baseline tendency when readers do not read any story.

We found that the relevancy of topics and moods within a user's browsing session is consistent with the flow construct (Csikszentmihalyi, 1997; Hoffman and Novak, 1996). According to this literature (Hoffman and Novak, 2009), consumers enter a state where they are completely immersed in a certain

5) For ease of interpretation, hereafter, the estimated coefficients in the <Table 6> are transformed in the exponential form, which indicates the relative increase or decrease in the impact of story topics on the headline topic preferences.

activity or interactions on the web, and this state gets more prevalent as the level of curiosity aroused increases from the richness and novelty of the content (Huang, 2003). Once they are in the state of flow, they present explorative and obsessive behaviors (e.g., exploiting all kinds of online recipes, engrossed with a certain type of web content; Novak et al., 2003), sometimes intensifying the preoccupation into an addiction to the habitual flow experience. In our study, the state of flow can be manifested in the way consumers explore news articles within a boundary of topical relevance and mood consistency. The phenomenon also conforms to findings on the maintenance of positive mood valence in consumer choice (e.g., Di Muro and Murray, 2012) and the pursuit of depth over breadth in the consumption of streamed video content (Schweidel and Moe, 2016).

One of the important components of our model is to address how the overall attractiveness of the website affects a user's exit decision (Eq. 6). The attractiveness measure is a function of the entire utilities of articles displayed on the choice set and a user's own tendency to browse an outside option (e.g., old articles, recommended articles). The model estimates a constant parameter to be $\hat{\delta}_0$ = -8.3 with the C.I. of [-8.5, -8.0] and an attractiveness parameter to be $\hat{\delta}_1$ =1.4 with the C.I. of [1.4, 1.4] in the deterministic term, V_{iqt}. Note that the higher value V_{iqt} has, the less likely a user is to exit. This indicates that if the overall utilities of all articles in the choice set are high, a user is prone to stay in, rather than exit, the website, which sounds plausible considering a news consumer's perspective.

How would each headline and story topic affect the exit decision? To answer this question, we increase the prevalence of each topic by 10% and investigate its lift of the exit probability based on the posterior estimates. To be specific, taking into consideration the nature of topic distribution in the covariates X_{iqtr} and Z_{iqtr} (in Eq. 1 and 4, respectively), we increase a focal topic by 0.1 (10%) and decrease other topics by [0.1/# of topics considered] for all articles in the dataset. Then, for the impact on the exit probability, we draw augmented values of V_{iqt} using the posterior parameter distribution and consequently the simulated exit probabilities (Eq. 6) for each article and average them out for comparison with the base exit probability (derived with the observed topic prevalence).

<Table 7> demonstrates the results of the topic effects on the exit probabilities, whereby the numbers indicate relative increase (>1.00) or decrease (<1.00). For example, P(exit)=1.022 for the politics/election topic implies that the exit probability increases by 2.2% when increasing politics/election topics more saliently compared to the baseline exit probability (i.e., the probability with the current setup). We find substantial variation across headline topics (6% decrease to 13% increase), whereas there was an almost ignorable impact by story topics (1% decrease to 1% increase). This finding is conceivable considering that users are more likely to exit or stay depending on the existence of interesting articles assessed by headlines, rather than the story they read before; though some stories may satiate users and discourage them to read further (perhaps some mundane articles, such as "Daily Life" stories, as captured in our data).

In summary, our news consumption model reveals that consumers largely make news consumption choices based on the 14 headline topics, and the 6 story topics from the articles that are previously viewed affect their subsequent article choice. In general, consumers are prone to choose more arousing and violent topics compared to mundane, every-day topics. Furthermore, after reading a story, consumers

<Table 7> Simulation Results of the Headline Topic Changes

Headline Topics	P(exit)	E(session length)
Politics/Election	1.022	0.969
Social issues/ Court rulings	1.007	0.990
Entertainment/ Sports	0.996	1.004
Website features	1.127	0.825
Crime	0.942	1.094
Natural disasters	0.973	1.042
Transportation accident	0.976	1.039
Gun violence	0.963	1.058
North Korea/ Mass murder	0.981	1.031
Family tragedy	0.988	1.018
ISIS/ Terrorism	0.987	1.021
Police brutality	0.981	1.031
Health/ Home	1.022	0.966
World news	1.026	0.961

tend to maintain the similar mood, topic, and regional orientation. Finally, the overall satisfaction engendered by the displayed articles and their preferences positively affects the user's retention to continue to read articles within a session. These results provide meaningful insights for news organizations in terms of which topics matter to visitors based on their most recent browsing behaviors. To investigate whether the content of headlines or stories has a bigger impact on visitor behavior and consequently advertising revenue, we next conduct a simulation study that allows exploring each topic's marginal effect on the exit probability and the expected length of the session.

4.4. Simulation Study to Understand the Impact of Topics on the Length of Session

Although the aforementioned model results capture the effects of topics on the choice of article, the update of topic preference influencing the next choice, and the exit decision, these results are in some sense too fragmented to draw comprehensive insights for publishers. In this section, we use the simulation study to understand the overall effect of topics (headlines and stories, separately) on the overall consumption of articles within a session. In other words, we estimate the impact of a 10% increase in each headline topic and story topic on the expected length of sessions, compared with the baseline values where we use the original values and estimated posterior parameter distributions. The specific simulation process involves augmenting each news consumption incidence (Eq. 1) and consequently updating the dynamic component for the subsequent consumption (Eq. 4) based on the chosen news' stories. The session is defined "to be over" when the augmented news choice is "exit" (Eq. 6), or the choice set observed in the dataset is exhausted. We explore the entire 14 headline topics and 6 story topics in the simulation study, for which results are shown in <Table 8>.

The impact of the headline topics, which directly

<Table 8> Simulation Results of the Story Topic Changes

Story topics	P(exit)	E(session length)
Health/ Science	0.999	0.966
Politics	0.997	0.968
International	1.003	0.962
Crime	1.001	0.962
Daily life	1.06	1.018
National Security	0.994	0.975

affect the choice of an article as well as the exit decision, captures consistent insights with the main model results. An increase in the topics that we earlier characterize as stimulating and appalling (e.g., all crime-related topics) has a positive effect on the session length. On the other hand, we find a decrease in the session length when those mundane and everyday-related headline topics (e.g., "*Politics/ Election,*" "*Social issues/ Court rulings,*" "*Website features,*" "*Health/ Home,*" "*World news*") are increased by 10%. The largest positive effect on session length is associated with increasing the topic "*Crime*" (9% increase in session length), whereas the most seemingly monotonous topic "*Website features*" shows the biggest negative impact on session length (18% decrease in session length).

On the other hand, the results of the impact of story topics have a different implication compared to those of headline topics. For instance, those stimulating stories such as "*Crime*" seem to result in only negative impact, whereas mundane stories seem to result in a positive effect alternately on the session length ("*Daily life*") or retention ("*Health/ Science,*" "*Politics,*" "*National Security*"). These findings are even more interesting if we consider the effect of the similar kinds of headline topics. For instance, when headlines are related to "*Crime,*" consumers tend to stay on the site and read more articles, yet when they read "*Crime*" stories, they are prone to leave the site and terminate the session earlier. It is possible that the headlines about those arousing stories may make the website more exciting and interesting. However, once a visitor consumes the content, he may become satiated with such arousing stories, which in turn negatively affects how long he will remain on the website and how many pages he will visit during the session. On the other hand, the headlines about the mundane stories may not look exciting initially. Once a visitor reads such an article, however, he may enter a flow state of positive mood and continuously select articles of similar moods ("*Daily life*") or cues from the stories previously viewed. We suspect that these everyday stories may engage the reader into the flow state with less resistance than what can be induced by the agitating stories such as "Crime." Contrary to the negative effects of the dominance of public affairs-related hard news topics on headlines, we find positive retention effects from such topics in the story content.

This relatively simple simulation allows us to measure the value of news content within the context of an individual's news consumption session. The separation of headline and story topics further provides holistic insights about dynamic contributions of topics depending on where they appear, which differs from the contemporary metrics used by news organizations in gauging content value by examining the performance of individual articles or aggregated

groups of content to assess site-wide engagement.

Ⅴ. Discussion

Despite the importance of understanding the dynamics that drive the consumption of information (Montgomery et al., 2004; Park and Park, 2016), little work has investigated the dynamic preference for news content expressed by individual consumers. With news being increasingly consumed through digital platforms in recent years, we take a visitor-centric approach to examining news consumption. Viewing individual stories as "products" and the content of those stories (summarized by latent topics) as "attributes," we borrow from the customer relationship management literature (Payne and Frow, 2005; Verhoef, 2003) to investigate how visitors' preferences evolve as they consume content and what impact the shifts in the content of stories may have on future consumption and consequently visitors' value to the organization. While prior research in this domain has mainly focused on instrumental factors (e.g., links, user-generated content), this is, to the best of our knowledge, the first study to understand news visitors' content consumption behaviors based on content characteristics, the key attribute that the news organizations deliver.

Furthermore, the current study sheds light on state dependence theory (Novak et al., 2003) and provides deeper insights about the direction of how the theory operates among news consumers. This study shows clear evidence that visitors' preferences for news content evolve based on the content they have consumed previously. In particular, our results suggest that visitors exhibit a degree of stickiness in the types of stories they read. Rather than finding visitors jumping between more somber content and lighthearted content, visitors are more likely to restrict their consumption of news to similar topics or topics with a similar tone. This is in line with prior research on media consumption (e.g., Schweidel and Moe, 2016), as well as research on the development of echo chambers and the spread of misinformation (e.g., Del Vicario et al., 2016; Schmidt et al., 2017). Visitors' inherent preferences not only drive their decisions about the content they choose to read, but also, as visitors consume more content, those choice preferences are reinforced. Thus, even with the availability of a diverse range of content on a platform, we see that visitors are more prone to seek out content similar to what they have already consumed.

Beyond furthering our understanding of the information that consumers choose to seek out, such insights are directly relevant to media organizations that generate revenue by embedding advertising in their content and delivering it to consumers. Our analysis reveals that increasing the prevalence of certain topics is expected to increase the length of visitors' sessions. Interestingly, we see that changes in the prevalence of topics in the headlines of articles has a larger impact than changes to the content of the stories. Although our findings suggest that increasing the amount of content focusing on certain topics will increase visitors' sessions, content providers such as news organizations and media platforms must consider the balance between the depth and breadth of the content they offer.

For news organizations, one solution is to tailor the user experience based on the content that visitors have already consumed. The current recommendation system news organizations often adopt is based on popular articles and articles in the same section that the visitor historically consumed. Our research enriches such a recommendation system by separating a recommendation system to attract

visitors and to retain news consumers longer on the website. For example, our results suggest that, on one hand, news organizations can populate a list of violent or sensational articles and advertise them on social media platforms to attract new visitors to the website; whereas, on the other hand, they can retain visitors longer on the website by suggesting articles on certain topics that reflect visitors' revealed topical interests and inherent readership mood.

5.1. Limitations and Future Research

Our research could be extended in a number of directions. We employ data from a single website, for example, whereas individuals may visit multiple websites to consume news. Data about news consumption from multiple websites would allow better understanding of how consumers seek out information. Such an analysis would reveal if individuals visit websites that offer different perspectives or if they restrict the information they consume to a single perspective.

A related issue would be to investigate the impact that news aggregators and social media platforms have on individuals' news consumption. Users of such platforms can be exposed to a summary of the news content without leaving the website, yet news organizations do not generate revenue directly from these sites. Recently subscription revenue has grown at existing news outlets as online news providers have increasingly adopted the paid subscription model. Data on both subscribers and non-subscribers would help identify whether the two groups portray divergent consumption patterns.

Finally, our data also do not contain information about news consumption on mobile devices, which is of interest to newsrooms that employ editorial data analytics on the platforms for incoming web traffic. Recently, several research has shown that consumers are governed by different mindset when they use mobile devices versus desktops (Grewal and Stephen, 2019). Although we only focus on desktop users to restrict potential confounding factors from multi-device usage, it would be interesting to investigate how news consumers behave differently across devices and the implications of these differences for news organizations.

<References>

[1] Ahn, D. H., Jin, S. A. A., and Ritterfeld, U. (2012), "Sad movies don't always make me cry" the cognitive and affective processes underpinning enjoyment of tragedy. *Journal of Media Psychology, 24*(1), 9-18.

[2] Bell, D. R., and Lattin, J. M. (1998). Shopping behavior and consumer preference for store price format: Why 'large basket' shoppers prefer EDLP. *Marketing Science, 17*(1), 66-88.

[3] Berger, J., and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192-205.

[4] Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., and Quattrociocchi, W. (2016). Users polarization on Facebook and Youtube. *PLoS ONE, 11*(8), e0159641.

[5] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022.

[6] Bucklin, R. E., and Sismeiro, C. (2003). A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research, 40*(3), 249-267.

[7] Büschken, J., and Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science, 35*(6), 953-975.

[8] Chae, I. Y., Schweidel, D. A., Evgeniou, T., and

Padmanabhan, V. (2017). Analyzing content consumption in a hybrid content environment. *Working Paper*.

[9] Cherubini, F., and Nielsen, R. K. (2016). *Editorial analytics: How news media are developing and using audience data and metrics*. Oxford: Reuters Institute for the Study of Journalism.

[10] Chiou, L., and Tucker, C. (2013). Paywalls and the demand for news. *Information Economics and Policy, 25*(2), 61-69.

[11] Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life* (1st ed.). New York, NY: Basic Books.

[12] Datta, H., Knox, G., and Bronnenberg, B. J. (2017). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science, 37*(1), 1-175.

[13] Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences, 113*(3), 554-559.

[14] Di Muro, F., and Murray, K. B. (2012). An arousal regulation explanation of mood effects on consumer choice. *Journal of Consumer Research, 39*(3), 574-584.

[15] Fader, P. S., and Hardie, B. G. S. (1996). Modeling consumer choice among SKUs. *Journal of Marketing Research, 33*(4), 442-452.

[16] Fader, P. S., and Lattin, J. M. (1993). Accounting for heterogeneity and nonstationarity in a cross-sectional model of consumer purchase behavior. *Marketing Science, 12*(3), 304-317.

[17] Farquhar, P. H., and Rao, V. R. (1976). A balance model for evaluating subsets of multiattributed items. *Management Science, 22*(5), 528-539.

[18] Gilbride, T. J., Allenby, G. M., and Brazell, J. D. (2006). Models for heterogeneous variable selection. *Journal of Marketing Research, 43*(3), 420-430.

[19] Gilbride, T. J., and Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science, 23*(3), 391-406.

[20] Godes, D., and Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. *Marketing Science, 31*(3), 448-473.

[21] Goldenberg, J. L., Pyszczynski, T., Johnson, K. D., Greenberg, J., and Solomon, S. (1999). The appeal of tragedy: A terror management perspective. *Media Psychology, 1*(4), 313-329.

[22] Grewal, L., and Stephen, A. T. (2019). In mobile we trust: The effects of mobile versus nonmobile reviews on consumer purchase intentions. *Journal of Marketing Research, 56*(5), 791-808.

[23] Guadagni, P. M., and Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2*(3), 203-238.

[24] Heckman, J. J. (1991). Identifying the hand of past: Distinguishing state dependence from heterogeneity. *The American Economic Review, 81*(2), 75-79.

[25] Hoffman, D. L., and Novak, T. P. (2009). Flow online: Lessons learned and future prospects. *Journal of Interactive Marketing, 23*(1), 23-34.

[26] Hoffman, D. L., and Novakm, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing, 60*(3), 50-68.

[27] Huang, M. H. (2003). Designing website attributes to induce experiential encounters. *Computers in Human Behavior, 19*(4), 425-442.

[28] Kahn, B. E., Kalwani, M. U., and Morrison, D. G. (1986). Measuring variety-seeking and reinforcement behaviors using panel data. *Journal of Marketing Research, 23*(2), 89-100.

[29] Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics, 15*(3), 310-327.

[30] Kim, J. H., Allenby, G. M., and Rossi, P. E. (2007). Product attributes and models of multiple discreteness. *Journal of Econometrics, 138*(1), 208-230.

[31] Kumar, V., Anand, B., Gupta, S., and Oberholzer-Gee, F. (2012). The New York Times paywall. *Harvard Business School Case 512-077*, February 2012.

[32] Lattin, J. M. (1987). A model of balanced choice

behavior. *Marketing Science, 6*(1), 48-65.

[33] Li, S., Sun, B., and Wilcox, R. T. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research, 42*(2), 233-239.

[34] Mares, M. Lo., and Cantor, J. (1992). Elderly viewers' responses to televised portrayals of old age. *Communication Research, 19*(4), 459-478.

[35] McAlister, L. (1982). A dynamic attribute satiation model of variety-seeking behavior. *Journal of Consumer Research, 9*(2), 141-150.

[36] Meloy, M. G. (2000). Mood-driven distortion of product information. *Journal of Consumer Research, 27*(3), 345-359.

[37] Mitchell, A., Holcomb, J., and Weisel, R. (2016). *State of the news media 2016*. Available at http://www.journalism.org/2016/06/15/state-of-the-news-media-2016/

[38] Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology, 13*(1-2), 29-39.

[39] Moe, W. W., and Fader, P. S. (2004). Dynamic conversion behavior at e-commerce sites. *Management Science, 50*(3), 326-335.

[40] Moe, W. W., and Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science, 31*(3), 372-386.

[41] Montgomery, A. L., Li, S., Srinivasan, K., and Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science, 23*(4), 579-595.

[42] Mullainathan, S., and Shleifer, A. (2005). The market for news. *The American Economic Review, 95*(4), 1031-1053.

[43] Newton, M. A., and Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological), 56*(1), 3-48.

[44] Novak, T. P., Hoffman, D. L., and Duhachek, A. (2003). The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology, 13*(1-2), 3-16.

[45] Park, C. H., and Park, Y. H. (2016). Investigating purchase conversion by uncovering online visit patterns. *Marketing Science, 35*(6), 894-914.

[46] Payne, A., and Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing, 69*(4), 167-176.

[47] Roos, J. M. T., Mela, C. F., and Shachar, R. (2015). The effect of links and excerpts on internet news consumption. *Working Paper*.

[48] Roy, R., Chintagunta, P. K., and Haldar, S. (1996). A framework for investigating habiits, "the hand of the past," and heterogeneity in dynamic brand choice. *Marketing Science, 15*(3), 280-299.

[49] Salton, G. (1971). *The SMART retrieval system-experiments in automatic document processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

[50] Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences, 114*(12), 3035-3039.

[51] Schweidel, D. A., and Moe, W. W. (2016). Binge watching and advertising. *Journal of Marketing, 80*(5), 1-19.

[52] Schweidel, D. A., Bradlow, E. T., and Fader, P. S. (2011). Portfolio dynamics for customers of a multiservice provider. *Management Science, 57*(3), 471-486.

[53] Simonov, A., and Rao, J. (2017). Demand for (un)biased news: government control in online news markets. *Working Paper*.

[54] Smith, G. (2017). Trump Bump for president's media archenemies eludes local papers. *Bloomberg*, Available at https://www.bloomberg.com/news/articles/2017-07-10/trump-bump-for-president-s-media-archenemies-eludes-local-papers

[55] Song, Y., Sahoo, N., and Ofek, E. (2016). When diversity becomes relevant-a multi-category utility model of consumer response to content recommendations. Working Paper.

[56] The New York Times. (2017). *The New York Times*

company reports 2017 second-quarter results. http://investors.nytco.com/press/press-releases/press-release-details/2017/The-New-York-Times-Company-Reports-2017-Second-Quarter-Results/default.aspx (Accessed on July 27 2021).

[57] Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing, 67*(4), 30-45.

[58] Xiang, Y., and Sarvary, M. (2007). News consumption and media bias. *Marketing Science, 26*(5), 611-628.

[59] Xiang, Y., and Soberman, D. (2014). Consumer favorites and the design of news. *Management Science, 60*(1), 188-205.

[60] Yildirim, P., Gal-Or, E., and Geylani, T. (2013). User-generated content and bias in news media. *Management Science, 59*(12), 2655-2666.

[61] Zhang, Y., Bradlow, E. T., and Small, D. S. (2015). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science, 34*(2), 195-208.

[62] Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2015). Emotional dynamics in the age of misinformation. *PloS ONE, 10*(9), e0138740.

<Appendix> Posterior Estimation Details

We use MCMC for posterior estimation. The sampling procedure is as follows:

i. Metropolis-Hastings algorithm to update $\Gamma_p|y, \boldsymbol{X}, \boldsymbol{Z}, U_0$

The conditional (S+1) vector parameter $\Gamma_p\ (p = 1, \dots, P)$ is as follows:

$$\pi(\Gamma_p|y, \boldsymbol{X}, \boldsymbol{Z}, U_0) \propto \prod_{i=1}^{N}\prod_{q=1}^{Q_i}\prod_{t=1}^{T_{iq}} \Pr(y_{iqt} = r)\, MVN(\mu_\gamma = 0, \Pi_\gamma = 100 I_{s+1})$$

We use a random-walk MH algorithm with a multivariate-t distribution: $\Gamma_p^{t+1} = \Gamma_p^t + \kappa_\Gamma$ where $\kappa_\Gamma \sim MVt(0, s_\Gamma T_\Gamma)$, and T_Γ is the empirical covariance from an extended burn-in period (Haario et al., 2005) so that the proposal density follows the approximate posterior covariance. The parameter s_Γ is adjusted such that the acceptance ratio belongs to 0.3~0.5.

ii. Metropolis-Hasting algorithm to update $\delta|y, \boldsymbol{X}, \boldsymbol{Z}, U_0, \Gamma$

The conditional vector parameter δ is as follows:

$$\pi(\delta|y, \boldsymbol{X}, \boldsymbol{Z}, \Gamma, U_0) \propto$$

$$\prod_{i=1}^{N}\prod_{q=1}^{Q_i}\prod_{t=1}^{T_{iq}} \Pr(y_{iqt} = exit)^{I(r=0)} (1 - \Pr(y_{iqt} = exit))^{I(r\neq 0)}\, MVN(\mu_\delta = 0, \Pi_\delta = 100 I_2)$$

We use a random-walk MH algorithm with a multivariate-t distribution: $\delta^{t+1} = \delta^t + \kappa_\delta$ where $\kappa_\delta \sim MVt(0, s_\delta T_\delta)$, and T_δ is the empirical covariance from an extended burn-in period (Haario et al., 2005) so that the proposal density follows the approximate posterior covariance. The parameter s_δ is adjusted such that the acceptance ratio belongs to 0.3~0.5.

iii. Metropolis-Hasting algorithm to update $U_{i0}|y, \boldsymbol{X}, \boldsymbol{Z}, \Gamma, \delta$

For each individual i, the individual-level baseline U_{i0} is drawn based on the conditional likelihood:

$$\pi(U_{i0}|y, \boldsymbol{X}, \boldsymbol{Z}, \Gamma, \delta) \propto$$

$$\prod_{q=1}^{Q_i}\prod_{t=1}^{T_{iq}} f(y_{it} = exit|\ \delta)^{I(r=0)}\{f(y_{it} \neq exit|\ \delta))f(y_{it} = r|\Gamma, \delta, U_{i0})\}^{I(r>0)} N(U_{i0}|\sigma^2)$$

For proposal density, we use a random walk MH algorithm with a normal distribution with variance Σ: $U_{i0}^{t+1} = U_{i0}^t + \kappa_u$ where $\kappa_u \sim N(0, \sigma^2)$

iv. Gibbs sampling to update σ^2

$$\pi(\sigma^2|U_0) \sim IW(n_k + \nu_0, D_0 + \textstyle\sum_{i=1}^{N} U_0^2)$$

◆ About the Authors ◆

Inyoung Chae

Inyoung Chae is an Assistant Professor of Marketing at Sungkyunkwan University.

Da Young Kim

Da Young Kim is a doctoral candidate at Emory University's Goizueta Business School.

Submitted: August 10, 2021; 1st Revision: October 25, 2021; Accepted: November 29 2021