Blazing the Trail: Considering Browsing Path Dependence in Online Service Response Strategy

Meihua Zuo[1] · Spyros Angelopoulos[2] · Zhouyang Liang[3] · Carol X. J. Ou[4]

Abstract

Competition on e-commerce platforms is becoming increasingly fierce, due to the ease of online searching for comparing products and services. We examine how the sequential browsing behavior of consumers can enable targeted marketing strategies on e-commerce platforms, by using clickstream data from one of the largest e-commerce platforms in Asia. We deploy duration analysis to i) explore how path dependence can better explain consumers' sequential browsing behavior in different product categories, and ii) characterize the sequential browsing behavior of heterogeneous consumer groups. The findings of our work showcase i) the high accuracy of using sequential browsing path dependence to explain consumer behavior, ii) the patterns of their behavioral intentions and iii) the spell of the behavior of heterogeneous consumer groups. Our findings provide nuanced implications for strategically managing branding, marketing, and customer relations on e-commerce platforms. We discuss the implications of our findings for both research and practice, and we delineate an agenda for future research on the topic.

Keywords E-commerce · Business Analytics · Consumer Behavior · Path Dependence · Duration Analysis

1 Introduction

The maturity of e-commerce has fundamentally changed the searching and purchasing behavior of consumers and has enabled them to compare products before purchasing (Bronnenberg et al., 2016). For instance, when purchasing a digital camera online, consumers engage on average in browsing 6.4 products from 3 companies (*ibid*), while when booking hotels online, they browse on average 2.3 companies, with large heterogeneity across groups (Chen & Yao, 2017). Prior studies have explored the relationship between purchasing intention and browsing behavior, such as depth (e.g., Johnson et al., 2004), frequency (e.g., Moe & Fader, 2004a, b), and duration of browsing (e.g., Sismeiro & Bucklin, 2004). We extend this line of research by considering consumers' browsing-path dependence.

Browsing-path dependence refers to a consumer's browsing behavior in one company and how it will affect this consumer's behavior in another. Considering the browsing path dependence can alleviate biased deviations between the interpreted purchasing intention and the actual behavior, which can guide a better understanding of consumers and enable decision support for service response strategies. Consumer service is defined as a useful, responsive service that responds quickly to consumer inquiries and returns or complaints during or after a purchase (Holloway & Beatty, 2008) and is one of the dimensions of e-service quality measurement (Behera et al., 2021; Wang et al., 2004). We specifically focus on service response during purchasing (Bhatnagar et al., 2017), due to its effect on consumer satisfaction, which affects purchasing behavior (Udo et al., 2010), as well as because prior studies have merely explored the topic from a consumer review perspective after purchasing (e.g., Proserpio & Zervas, 2017).

✉ Spyros Angelopoulos
  spyros.angelopoulos@durham.ac.uk

  Meihua Zuo
  zuomeihua123@126.com

  Zhouyang Liang
  liangzy@gdut.edu.cn

  Carol X. J. Ou
  carol.ou@uvt.nl

1 Huizhou University, Huizhou, China

2 Durham University Business School, Durham University, Durham, UK

3 Guangdong University of Technology, Guangzhou, China

4 Tilburg University, Tilburg, Netherlands

Compared to the attempts in the extant literature, the browsing path dependence could demonstrate more accurately the real-time purchasing intentions of consumers, thereby enabling decision support for e-commerce platforms to formulate dynamic service response strategies. To study the browsing path dependence of consumers, we use clickstream data that showcase their browsing behavior in a time-variant and objective way (Balan & Mathew, 2021; Bucklin & Sismeiro, 2009; Kim et al., 2011). Clickstream data have recently become a rich source of information for both researchers and practitioners to unearth the potential interests and preferences of heterogeneous consumers. We use a dataset from one of the largest e-commerce platforms in Asia capturing the time-variant browsing behavior of consumers for 11 weeks (77 days). We separately model the browsing path dependencies for overall consumers and heterogeneous consumer groups through duration analysis (Bhatnagar et al., 2017).

Our findings demonstrate that the actual behavior of consumers can be accurately described by considering the dependencies of the sequential browsing paths. We show that the path dependence of consumers decreases with the increase in browsing times. In practice, this means that companies need to provide a timely service response to first-time consumers. We also find that the behavior patterns of heterogeneous consumer groups are quite different, and the degree of such a difference is related to the position that a certain company holds in the market.

The rest of our paper is structured as follows. In the next section, we provide an overview of the literature. After this, the dataset and methodology of the study are discussed. The following section describes the results of our study. Next, these results are discussed, and their implications are also considered. Finally, a brief conclusion is provided along with limitations, as well as an agenda for future research on the topic.

2 Background

Clickstream data contain *inter alia* consumers' i) browsing path, ii) sequential purchase, iii) website entrance, and iv) other associated information (Bhatnagar et al., 2017; Montgomery, 2001). Browsing-path information differs from the simple sequential transaction information as it can record the complete trajectory of consumers in the online shopping process (Hui et al., 2009), and it is more likely to better explain consumers' interests and purchasing intentions. Prior studies have explored the purchasing intentions of consumers based on the characteristics of their online browsing behavior. For instance, Moe and Fader (2004a, b) explored the relationship between consumers' purchasing intentions and frequency of browses and showed that those who browse a website more frequently have a higher purchasing intention. Similarly, when consumers' purchasing process is decomposed, the longer consumers' browse a specific product page, the higher their purchasing intention is (e.g., Sismeiro & Bucklin, 2004). Such studies, therefore, provide a foundation to examine clickstream data and consumers' purchasing intention.

Most studies on online browsing behavior have investigated at the aggregate level, that is, they have studied the cumulative browsing behavior across all consumers. However, many studies have found that heterogeneous consumer groups have large differences in learning ability and consumption experience, resulting in large differences in browsing behavior (Bhatnagar et al., 2017; Johnson et al., 2003). In addition, Sismeiro and Bucklin (2004) found that the results of analysis using aggregate-level browsing data are often misleading. Therefore, we use the overall consumer behavior to illustrate the importance of browsing path-dependent information. We further explain the heterogeneity in consumers' browsing behavior by using behavioral information of heterogeneous consumer groups.

We argue that browsing-path dependence is critical for explaining consumers' behavior, especially the purchasing behavior of complex products. For example, a consumer may acquire product and company information from the webpage of company A and then purchase directly from another company B. If sequential browsing information is not integrated into the analysis, then consumers' behavior is calculated based only on partial information and the results can be biased (Park & Fader, 2004). Prior studies have demonstrated that integrating consumers' sequential path information can yield a better explanation. For instance, Park and Fader (2004) explained consumers' behavior on the website of one company based on the browsing-path information associated with frequency and timing when consumers switch to the websites of other companies. Karimi (2021) used information on consumers' cross-visit behavior between comparison and retailer websites to explain consumers' holistic behavior.

Prior studies have considered consumers' browsing-path behavior on multiple websites (e.g., Karimi, 2021; Park, 2017). We argue that it is equivalently, if not more, important to consider consumers' browsing-path behavior on multiple companies on the same website, since competition intensity on one platform is greater than that on different websites (Bronnenberg et al., 2016). In addition, explaining consumers' browsing path intentions can provide decision-making reference for timely service responses, and enhance companies' reputation, increasing their competition intensity. Thus, we deploy duration analysis to analyze the browsing-path dependence of related companies to explain the spell and browsing patterns of heterogeneous consumers.

3 Methodology

3.1 Data

We acquired data from one of the most popular e-commerce platforms in Asia, which sells millions of high-quality products from tens of thousands of companies in 12 product categories, which include: electrical appliances, digital cameras, computers, mobile phones, home furniture, clothes, shoes, luggage, cosmetics, mother and baby products, books, and food. To increase the practical implications of our findings, we adopted the product categorization of the focal e-commerce platform without further interference. We collected information about consumers' time-variant online browsing behavior for a total of 11 weeks (77 days) from this platform. The dataset contains consumers' sequential browsing paths and demographics, as well as information about the companies they browsed. The dataset was constructed by merging three databases: i) consumers' sequential browsing path, covering the time when the browsing behavior occurred and the name of the company being browsed; ii) attributes of the browsed companies, including the name and category the company belongs to; and iii) consumers' demographic information, including gender, membership level, and age. To ensure confidentiality, all information about consumers and companies is pseudonymized via numerical coding.

Since consumers' browsing behavior can vary based on the category of product, to verify the effectiveness of path dependence information, we follow Park and Fader (2004) who use consumers' browsing behavior data in two different categories of products. We present the summary statistics of consumers' browsing behavior in these two product categories in Table 1. For the convenience of the subsequent presentation, we named these two product categories 1 and 2. According to Table 1, the average number of times consumers repurchase products in category 2 is twice that of category 1, while the number of browses for category 1 products is more than twice that of category 2. Different categories of products have different attribute factors and consumption frequencies, so there is a difference in the number of consumers' browses and repurchases (e.g., Zuo et al., 2019). Therefore, according to consumers' browsing

and purchasing information, we can see these categories are different.

According to the preliminary statistics of the dataset, there are 66 companies in product category 1 and 127 in category 2. Due to the number of companies in each product category, we need to conduct company screening to show the dependence of consumers on the browsing paths of related companies in detail. Our preliminary statistics on product sales reveal that the cumulative volume of the top 10 companies in each product category accounts for about 90% of the entire market share. That is, the bottom 90% of companies in each product category have extremely low market shares, and the number of consumers with co-browsing behavior is sparse, making the results insignificant. Subsequently, we randomly combined the top 10 companies in each product category into two pairs and counted the number of consumers who co-browsed them. According to these results, we select the two companies with the most consumers, and we name these two companies in the first product category as A and B, and companies in the second product category as C and D. In these product categories, the number of consumers with browsing path records for at least one of the two related companies was 2,634 and 1,169, and the number of browsing records was 22,349 and 15,042, respectively. To study the impact of consumers' current browsing path information about related companies on subsequent browsing behavior, we exclude consumers who did not visit both companies at the same time. The number of consumers who browse two companies separately and both in different categories is shown in Table 2: 1,610 out of 2,634 consumers (61%) in the first category and 785 out of 1,169 consumers (67%) in the second one.

In line with prior studies (Moe & Fader, 2004a, b; Park & Fader, 2004), we also adopt the calendar day as a spell, which means if a consumer i browses a specific company multiple times on the same day, that company should be coded as browsed for that calendar day. For example, assuming consumer i first browses company A three times and then B for two times; then we code this consumer's browsing behavior as browsing both A and B that day.

To study the heterogeneity of consumers' browsing paths of related companies in the same product category, we classify consumers according to available demographic information including age (3 types), gender (2 types), and membership

Table 1 Statistics of consumers' behavior characteristics in two product categories

	No. of consumers with repurchase	Total No. of repurchase	No. of browses		
			Mean	Min	Max
Category 1	152	172	95.55	0	1182
Category 2	541	1029	40.28	0	744

Table 2 Statistics of consumers browsing two related companies

Category 1	Consumers	Browses	Category 2	Consumers	Browses
Company A	493	2,240	Company C	249	1,952
Company B	531	2,841	Company D	135	1,143
Both	1,610	17,268	Both	785	11,947
Total	2,634	22,349	Total	1,169	15,042

level (2 levels). Accordingly, consumers can be divided into 12 (3*2*2) types (see Table 3).

3.2 Data Analysis

Duration analysis is typically considered between two companies, but it can analyze path dependence among multiple ones (Park & Fader, 2004). We first present a simple bivariate distribution model that assumes consumers' browsing path among related companies is completely independent. Then we establish a Farlie-Gumbel-Morgenstern family model to show the dependence of the browsing paths (Chintagunta & Haldar, 1998; Park & Fader, 2004).

3.3 Benchmarking Models

Let us assume a random continuous variable T indicating that an event has been occurring until that time. A specific value of T_{ijl} is represented by t_{ijl}, in which both T_{ijl} and t_{ijl} are greater than or equal to 0, where i stands for different consumers, j represents different companies, and l represents different product categories. The possible values of j are 1 or 2, indicating two related companies in the same product category. Assume that the probability density and cumulative distribution function of t_{ijl} are $f(t_{ijl})$ and $F(t_{ijl})$ respectively, where $F(t_{ijl})$ refers to the failure function. The probability that the consumer browsing interval exceeds t_{ijl}, namely the univariate survival function, is defined in formula (1):

$$S(t_{ijl}) = P(T_{ijl} > t_{ijl}) = 1 - F(t_{ijl}), \tag{1}$$

Table 3 Consumer characteristics of browsing related companies in product category 1

Age	Gender	Membership level	No. of consumers (2,629)
≥ 35 (1)	Male (0)	non-VIP (1)	164
		VIP (2)	153
	Female (1)	non-VIP (1)	115
		VIP (2)	63
36–45 (2)	Male (0)	non-VIP (1)	361
		VIP (2)	555
	Female (1)	non-VIP (1)	224
		VIP (2)	183
46 ≥ (3)	Male (0)	non-VIP (1)	249
		VIP (2)	280
	Female (1)	non-VIP (1)	193
		VIP (2)	89

According to formula (1), the survival function $S(t_{ijl})$ is the inverse of $F(t_{ijl})$. Since $F(t_{ijl})$ is monotonically increasing, $S(t_{ijl})$ is monotonically decreasing. Therefore, if consumer i did not browse company j in product category l until time t_{ijl}, the hazard function of a specific instantaneous browsing probability can be expressed by formula (2):

$$h(t_{ijl}) = \frac{f(t_{ijl})}{S(t_{ijl})} = \frac{f(t_{ijl})}{1 - F(t_{ijl})}, \tag{2}$$

When considering the specific consumer's browsing paths among the related companies are independent, the joint cumulative distribution function is defined by formula (3):

$$F_{12}(t_{1_l}, t_{2_l}) = F_1(t_{1_l}) F_2(t_{2_l}), \tag{3}$$

where the function $F_{12}(t_{1_l}, t_{2_l})$ represents the joint cumulative distribution function of browsing paths between the company 1 and 2 in product category l, while $F_1(t_{1_l})$ and $F_2(t_{2_l})$ represent the marginal cumulative distribution functions, respectively.

3.4 Proposed Association Models

The most common way to express the correlation between two independent univariate distribution functions is to combine the two random variables with a third term. For instance, Farlie (1960) has proposed a common bivariate distribution function, as defined in formula (4):

$$F_{12}(t_{1_l}, t_{2_l}) = F_1(t_{1_l}) F_2(t_{2_l}) \{1 + \beta[1 - F_1(t_{1_l})][1 - F_2(t_{2_l})]\}, \\ -1 \leq \beta \leq 1 \tag{4}$$

The meaning of parameters in formula (4) is the same as formula (3), where the parameter β indicates the degree of dependence of consumers' browsing paths among related companies. According to Chintagunta and Haldar (1998), β can also identify the relationship between two companies. When β is negative, these companies have a potential complementary relationship, while when β is positive, they have a potential substitutional relationship (Park & Fader, 2004). If β is 0 or not significant, the browsing paths of these two companies are not relevant. The density function of the bivariate function in formula (4) is then expressed in formulas (5), (6), (7) and (8):

$$f_{12}(t_{1_l}, t_{2_l}) = \frac{\partial^2 F_{12}(t_{1_l}, t_{2_l})}{\partial t_{1_l} \partial t_{2_l}}, \\ = f_1(t_{1_l}) f_2(t_{2_l}) \{1 + \beta[2S_1(t_{1_l}) - 1][2S_2(t_{2_l}) - 1]\} \tag{5}$$

when consumer i browses both companies in the sequential browsing path, the browsing behavior to the likelihood

function is captured in formula (5), following Chintagunta and Haldar (1998). If a consumer only browses the first company, the browsing behavior is captured in formula (6), following Johnson and Kott (1975):

$$f_1(t_{1_l}, t_{2_l}) = f_1(t_{1_l})F_2(t_{2_l})\{1 + \beta[1 - 2F_1(t_{1_l})][1 - F_2(t_{2_l})]\}, \tag{6}$$

Similarly, when the consumer only browses the second company, the browsing behavior in the likelihood function is captured in formula (7):

$$f_2(t_{1_l}, t_{2_l}) = F_1(t_{1_l})f_2(t_{2_l})\{1 + \beta[1 - F_1(t_{1_l})][1 - 2F_2(t_{2_l})]\}, \tag{7}$$

When the consumer browses neither of these companies, the browsing behavior in the likelihood function is shown in formula (8):

$$S_{12}(t_{1_l}, t_{2_l}) = S_1(t_{1_l})S_2(t_{2_l})[1 + \beta F_1(t_{1_l})F_2(t_{2_l})], \tag{8}$$

3.5 Right Censoring of Association Models

A right censoring problem typically exists in consumers' sequential association browsing paths, meaning that when the pre-defined observation period (namely *spell*) ends, consumer i has no browsing behavior within the spell (Bhatnagar et al., 2017). As illustrated in Fig. 1, we set the starting observation time as 0, and the corresponding browsing behavior is also 0. The period from 0 to t_1 is spell 1. According to the illustrative example in Fig. 1, consumer i only browses company A in spell 1. In spell 3, this consumer does not browse any company. To show the continuity of the consumer's browsing behavior, we use right censoring to deal with the consumer's browsing behavior in the third spell. To address the potential right censoring problem, a dummy variable d_{ijl} is introduced in the likelihood function to indicate whether consumer i has the browsing behavior in the respective spell. When a consumer is interested in two related companies A and B, there are four possibilities for their browsing paths in any one spell: i) browsing only A; ii) browsing only B; iii) browsing both; and iv) browsing neither of these. Therefore, when consumer i has browsing behavior during this spell, then the contribution of the sequential browsing path to the likelihood function is shown in the above formulas (5), (6), (7); otherwise, the function is shown in formula (8).

3.6 Synthetic Likelihood Function of Association Models

Considering the four browsing paths during the observation period, the likelihood function for all browsing which is labeled in formula (9).

$$L = \prod_{\zeta=1}^{M}[f_{12}(t_{1_l}, t_{2_l})]^{\delta_{12\zeta}}[f_1(t_{1_l}, t_{2_l})]^{\delta_{10\zeta}}[f_2(t_{1_l}, t_{2_l})]^{\delta_{02\zeta}}[S_{12}(t_{1_l}, t_{2_l})]^{d_{ijl}}, \tag{9}$$

where M represents the number of completed spells, when a certain spell ζ ends, the consumer simultaneously browses the two companies in product category l, then the value of $\delta_{12\zeta}$ is equal to 1; otherwise, $\delta_{12\zeta}$ is equal to zero. If the consumer only browses the first company, then the value of $\delta_{10\zeta}$ is equal to 1; otherwise, $\delta_{10\zeta}$ is equal to zero. If the consumer only browses the second company, the value of $\delta_{02\zeta}$ is equal to 1; otherwise, $\delta_{02\zeta}$ is equal to zero. The last term of function is used to explain the right censoring problem, meanwhile, each browsing path can only be one of the above four types, which means the sum of $\delta_{12\zeta}$, $\delta_{10\zeta}$, $\delta_{02\zeta}$ and d_{ijl} must equal to 1.

3.7 Distribution Assumptions of the Density Function

Before conducting the parameter estimation, the distribution of the density function $f(t_{ijl})$ of the continuous random variable t_{ijl} needs to be assumed, satisfying the non-monotonousness assumption (Chintagunta & Haldar, 1998). The non-monotonic density functions commonly used in duration analysis include *exponential distribution* (Park & Fader, 2004), *Weibull*, and *Log-logistic* distribution (Bhatnagar et al., 2017; Chintagunta & Haldar, 1998). Among them, the exponential density function is the earliest hypothetical form. However, since the hazard function $h(t_{ijl})$ corresponding to the exponential density function is constant, the probability of consumer i browsing the two companies is unrelated to the length of the browsing interval. The hazard function of the exponential is memoryless, which differs from the consumer i in practice. Therefore, the exponential density function is extended with only one parameter to Log-logistic, and Weibull distribution with two parameters. After the preliminary processing of the data, we find that the density function of Log-logistic is better than Weibull distribution. Therefore, we assume that the density function $f(t_{ijl})$ of the sequential browsing path of consumer i follows a *Log-logistic* distribution, and the hazard function shown in formula (10):

Fig. 1 The Browsing Path of Consumer i
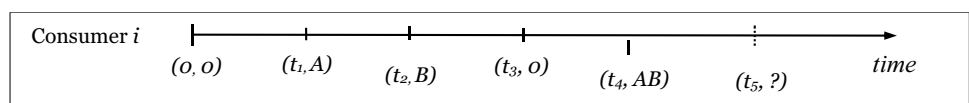
$$h(t_{ijl}) = \frac{\alpha \gamma t_{ijl}^{\gamma-1}}{1 + \alpha t_{ijl}^{\gamma}}, \tag{10}$$

α in formula (10) represents the scale parameter, with other parameters unchanged, a larger value of α indicates that consumer i is more likely to browse this company again. Here, γ represents the distribution of the hazard function. When $\gamma > 1$, the distribution of the hazard function is a single peak, which means that when consumer i enters a company store, the probability of browsing increases due to the need to first receive information about the company. However, once a certain time is reached, the browsing possibility starts declining, which means that when consumer i has a certain understanding of a company, they will concentrate on understanding a certain product. Therefore, the corresponding browsing probability will be reduced. However, when $\gamma \leq 1$, the change law of the hazard function is monotonically decreasing, that is, as time increases, the possibility of browsing each company will gradually decrease. This demonstrates that the consumer i is familiar with the company and has a clear shopping goal.

3.8 Bayesian Estimation Process of Parameters

To explore the uncertainty and heterogeneity of the estimated parameters, we use the Bayesian method for their estimation (Chu et al., 2017; Manouchehri et al., 2020). According to the Bayesian principle, the joint posterior distribution of each parameter can be expressed by formula (11):

$$p(\alpha_{i1l}, \gamma_{i1l}, \alpha_{i2l}, \gamma_{i2l}, \beta_{il}|t_{i1l}, t_{i2l})$$
$$\propto L(t_{i1l}, t_{i2l}|\alpha_{i1l}, \gamma_{i1l}, \alpha_{i2l}, \beta_{il})p(\alpha_{i1l})p(\gamma_{i1l})p(\alpha_{i2l})p(\gamma_{i2l})p(\beta_{il})'$$
$$\tag{11}$$

The likelihood function $L(\bullet)$ on the right side of formula (11) is calculated by the formula (9), and $p(\bullet)$ is the prior distribution of each parameter. If no prior information exists for each parameter, then the prior can be set to the suggestion of weakly informative priors (Gelman et al., 2013). Therefore, in the *Markov chain Monte Carlo* (MCMC) sampling process, the prior distribution settings for each parameter are shown in formula (12):

$$\begin{aligned}
\alpha_{i1l} &\sim HalfNormal(\sigma_{\alpha_{i1l}}^2) \\
\alpha_{i2l} &\sim HalfNormal(\sigma_{\alpha_{i2l}}^2) \\
\gamma_{i1l} &\sim HalfNormal(\sigma_{\gamma_{i1l}}^2) \\
\gamma_{i2l} &\sim HalfNormal(\sigma_{\gamma_{i2l}}^2) \\
\beta_{il} &\sim Uniform(-1,1)
\end{aligned} \tag{12}$$

3.9 Markov Chain-Monte Carlo Parameter Estimation

We use the No-U-Turn Sampler (NUTS) algorithm to sample and estimate the above parameters, which is an efficient parameter optimization method (Hoffman & Gelman, 2014). In each sampling, a recursive algorithm generates posterior information of candidate parameter sets. If the constraint of *non-U-type rotation* is satisfied, sub-tree updating parameters are constructed continuously. Otherwise, record the optimal parameter set of this sampling and start the next sampling until sufficient samples are obtained. The algorithm fully optimizes the parameters in each sampling, avoids redundant sampling due to random walk which, and improves the efficiency of parameter optimization. We run two independent MCMC to sample the parameters. Each MCMC was sampled 10,000 times and keeps a sample every 5 times. Hence, the number of samples used as a posterior inference of our model is 20,000.

4 Results

4.1 Estimation of all Parameters

We summarize the estimated results of parameters in Table 4: *Alpha* represents the scale parameter and *Gamma* represents the shape parameter of each company in the two product categories. *Beta* indicates the degree of dependence between consumers' browsing paths in related companies. We use Watanabe-Akaike Information Criterion (WAIC) (Vehtari et al., 2017; Watanabe, 2010) as the indicator that evaluates the model's fit effect, in which the smaller the value, the better the fit of the model. We use Mean (s.d.) to represent the average (standard deviation) of the estimated parameters. The interval value of 2.5–97.5% represents the 95% confidence level of the estimated parameters. If 0 is not in this interval, the parameter to be estimated is significant.

The parameter estimation and WAIC values indicate that, first, when comparing WAIC values of the benchmarking model and the proposed association model in these two product categories, all WAIC values of the proposed association models are smaller than the corresponding benchmarking model. This suggests, that regardless of category, the model-fitting effect of the proposed association model with browsing-path dependence is better than that of the benchmarking one. Second, all parameters are significant at the 95% confidence interval corresponding to the estimated value of each parameter and the judgment criteria of the Bayesian estimation method. Third, the maximum value of the average standard deviation is 0.017, and most of the other values are within 0.010, indicating that the distribution of the estimated values of each parameter is relatively concentrated. Fourth, the significant positive beta indicates that the related companies in the two categories are competing (Park & Fader, 2004).

Table 4 Estimation of parameters

Parameter	Category 1: Companies A & B				Category 2: Companies C & D			
	Benchmarking Model		Proposed Association Model		Benchmarking Model		Proposed Association Model	
	Mean (s.d.)	(2.5–97.5)%	Mean (s.d.)	(2.5–97.5)%	Mean (s.d.)	(2.5–97.5)%	Mean (s.d.)	(2.5–97.5)%
Alpha1 (scale- company A/C)	0.233* (0.003)	(0.22,0.23)	0.208* (0.003)	(0.20,0.21)	0.233* (0.005)	(0.23,0.24)	0.168* (0.006)	(0.41,0.43)
Gamma1 (shape- company A/C)	1.703* (0.004)	(0.70,0.71)	0.717* (0.004)	(0.71,0.73)	1.365* (0.012)	(1.34,1.39)	0.836* (0.011)	(1.28,1.32)
Alpha2 (scale- company B/D)	0.318* (0.004)	(0.31,0.33)	0.276* (0.004)	(0.27,0.28)	0.418* (0.007)	(0.40,0.43)	0.256* (0.008)	(0.64,0.68)
Gamma2 (shape- company B/D)	1.629* (0.004)	(0.62,0.64)	0.652* (0.004)	(0.64,0.66)	1.455* (0.015)	(1.43,1.48)	0.767* (0.010)	(1.39,1.44)
Beta (browsing-path dependence)	nil	nil	0.999* (0.001)	(0.99,1.00)	nil	nil	0.981* (0.017)	(0.99,1.00)
WAIC	224149.35		221040.86		143321.45		140962.02	

*: Significant at 95% confidence interval

Table 5 Statistics of consumers browsing two related companies

Category 1	Consumers	Browses	Category 2	Consumers	Browses
Company E	296	987	Company G	70	333
Company F	885	5,467	Company H	236	1,627
Both	664	5,618	Both	412	4,180
Total	1,845	12,072	Total	718	6,140

To test the robustness of the results in Table 4, we randomly select two of the other top 8 companies in sales volume. We name the companies in the first product category as *E* and *F*, and in the second category as *G* and *H*. In these categories, the number of consumers with browsing path records for at least one of the two related companies was 1,845 and 718 respectively, while the browsing records for these categories were 12,072 and 6,140, respectively. The number of unique consumers who browse two companies separately in different product categories is shown in Table 5, which shows that 664 out of 1,845 consumers (36%) in the first product category and 412 out of 718 consumers (57%) in the second one. We summarize the results of the parameters and robustness test in Table 6.

The parameter estimation and WAIC values indicate the following results, which demonstrate the robustness of the results in Table 4:

Table 6 Estimation of parameters

Parameter	Category 1: Companies E & F				Category 2: Companies G & H			
	Benchmarking Model		Association Model		Benchmarking Model		Association Model	
	Mean (s.d.)	(2.5–97.5)%	Mean (s.d.)	(2.5–97.5)%	Mean (s.d.)	(2.5–97.5)%	Mean (s.d.)	(2.5–97.5)%
Alpha1 (scale-company E/G)	0.25* (0.006)	(0.23,0.26)	0.229* (0.006)	(0.22,0.24)	0.178* (0.006)	(0.17,0.19)	0.168* (0.006)	(0.16,0.18)
Gamma1 (shape-company E/G)	0.71* (0.008)	(0.69,0.72)	0.723* (0.008)	(0.71,0.74)	0.83* (0.010)	(0.81,0.85)	0.836* (0.011)	(0.82,0.86)
Alpha2 (scale-company F/H)	0.20* (0.006)	(0.19,0.21)	0.171* (0.005)	(0.16,0.18)	0.269* (0.008)	(0.25,0.28)	0.256* (0.008)	(0.24,0.27)
Gamma2 (shape-company F/H)	0.74* (0.009)	(0.72,0.76)	0.761* (0.009)	(0.74,0.78)	0.759* (0.010)	(0.74,0.78)	0.767* (0.010)	(0.75,0.79)
Beta (browsing-path dependence)	nil	nil	0.996* (0.004)	(0.99,1.00)	nil	nil	0.981* (0.017)	(0.95,1.00)
WAIC	57850.20		56990.50		43448.00		42955.90	

*: Significant at 95% confidence interval

1. when comparing WAIC values of the benchmarking model and the proposed association model in these two product categories, all WAIC values of the proposed association models are smaller than the corresponding benchmarking model. This suggests that the model-fitting effect of the proposed association model with browsing-path dependence is better than that of the benchmarking model, regardless of product category.

2. all parameters are significant at the 95% confidence interval corresponding to the estimated value of each parameter and the judgment criteria of Bayesian estimation.

3. the maximum value of the average standard deviation is 0.017, and most of the other values are within 0.010, indicating that the distribution of the estimated values of each parameter is relatively concentrated.

4. the significant positive beta indicates that the related companies in each of the two product categories are in a competitive relationship (Park & Fader, 2004).

We illustrate consumers' dependence on their paths when browsing related companies in Fig. 2, where the x-axis and y-axis denote the distribution of browsing time for each of the two companies, and *deltat* represents the time interval between the current and the last browsing behavior. We use the Kernel Density Estimation to fit the correlation of browsing time. According to Fig. 2, the path dependence of most consumers occurs in the first four calendar days. Among them, the dependence on the browsing path of related companies is the highest on the first calendar day, and then it gradually decreases, because when consumers browse related companies for the first time, they need to compare them. With the increase in the number of browses, the information of related companies is gradually familiar, so no comparison is made, which leads to a gradual decrease in the degree of path dependence.

4.2 Parameter Estimations Considering Consumer Heterogeneity

The literature suggests that the browsing paths, frequency of browsing, and the time spent on each webpage for different types of consumers can be distinct (Su & Chen, 2015). Our findings show that consumers significantly differ in their online browsing strategies. We, therefore, categorize consumers into 12 types based on their demographics (i.e., 2 gender categories * 3 age types * 2 membership levels). To analyze the heterogeneity of consumer browsing behavior, we choose the consumer sequential browsing-path information of the first product category as the illustration. Each type of consumer is labeled with a three-digit number: the first digit indicates the age group (1, 2, or 3), the second digit indicates the consumer's gender (0 or 1), and the third
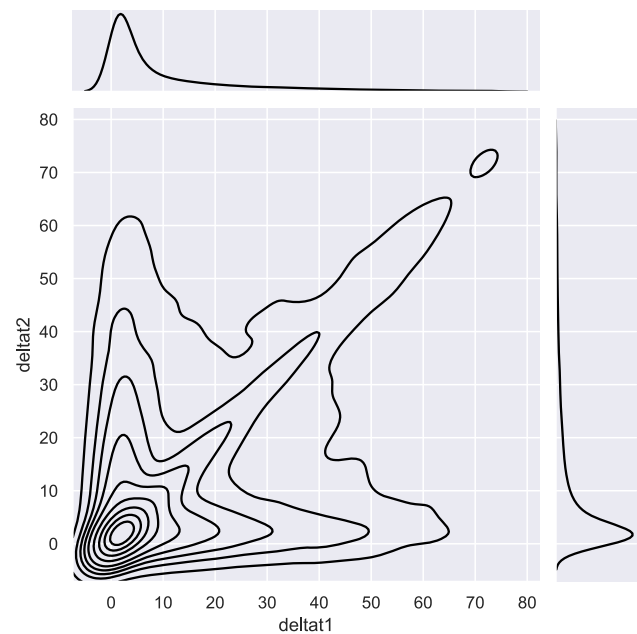

Fig. 2 The effect of path dependence. Note: The axes represent time interval between current and last browsing in days. The side graphs show distribution of consumers' browsing time in the two companies. The main graph represents the frequency of consumers' browsing path dependence on both companies, and the denser line indicates higher browsing path dependence

digit indicates the consumer's membership level (1 or 2). For instance, the type with code 102 refers to VIPs, males, and under 35 years old. In Table 7 we demonstrate the browsing path characteristics of the 12 consumer groups in related companies.

Our analyses demonstrate that Gamma values have two types of distributions: i) both Gamma values are greater than 1 for groups 101, 111, 201, 202, 211, 301, and 302; ii) both Gamma values are less than 1 for groups 102, 112, 212, 311 and 312. We compare the browsing of heterogeneous consumer groups according to the above distribution of *Gamma* values by incorporating the following two consumer groups in the detailed analysis: 112 in type (1), and 211 in type (2). We present the Hazard Function of these two consumer groups in Fig. 3, where the three-digit number refers to the focal consumer group. Because each consumer group browses two related companies at the same time, there are two curves of browsing path characteristics for each consumer group. Among them, *Alpha1* and *Gamma1* represent the scale and shape parameters of company A respectively, while *Alpha2* and *Gamma2* represent the scale and shape parameters of company B respectively.

Several observations can be obtained from these results. Consumer groups such as 101, 111, 201, 202, 211, 301, and 302 have both Gamma values greater than 1, suggesting that these consumer groups are different from all other

Table 7 Consumers' browsing path characteristics of companies A & B

Parameter	101 Mean (s.d.)	102 Mean (s.d.)	111 Mean (s.d.)	112 Mean (s.d.)	201 Mean (s.d.)	202 Mean (s.d.)	211 Mean (s.d.)	212 Mean (s.d.)	301 Mean (s.d.)	302 Mean (s.d.)	311 Mean (s.d.)	312 Mean (s.d.)
Alpha1 (scale-company A)	0.36 (0.02)	0.19 (0.01)	0.34 (0.03)	0.19 (0.02)	0.38 (0.02)	0.37 (0.01)	0.36 (0.02)	0.19 (0.01)	0.40 (0.02)	0.33 (0.02)	0.17 (0.01)	0.22 (0.02)
Gamma1 (shape-company A)	1.29 (0.04)	0.80 (0.02)	1.22 (0.04)	0.76 (0.03)	1.41 (0.03)	1.37 (0.02)	1.34 (0.03)	0.75 (0.02)	1.38 (0.03)	1.21 (0.03)	0.75 (0.02)	0.76 (0.02)
Alpha2 (scale-company B)	0.51 (0.03)	0.27 (0.02)	0.45 (0.04)	0.27 (0.02)	0.53 (0.02)	0.46 (0.01)	0.56 (0.03)	0.30 (0.02)	0.61 (0.03)	0.45 (0.02)	0.30 (0.02)	0.34 (0.03)
Gamma2 (shape-company B)	1.27 (0.04)	0.69 (0.02)	1.26 (0.05)	0.65 (0.03)	1.41 (0.03)	1.36 (0.02)	1.22 (0.03)	0.62 (0.02)	1.40 (0.04)	1.19 (0.03)	0.62 (0.02)	0.63 (0.02)
Beta (Browsing-Path Dependence)	0.99 (0.02)	0.97 (0.03)	0.98 (0.02)	0.98 (0.03)	0.99 (0.00)	0.99 (0.00)	0.99 (0.01)	0.98 (0.02)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.97 (0.03)

Fig. 3 Hazard function of different consumer groups
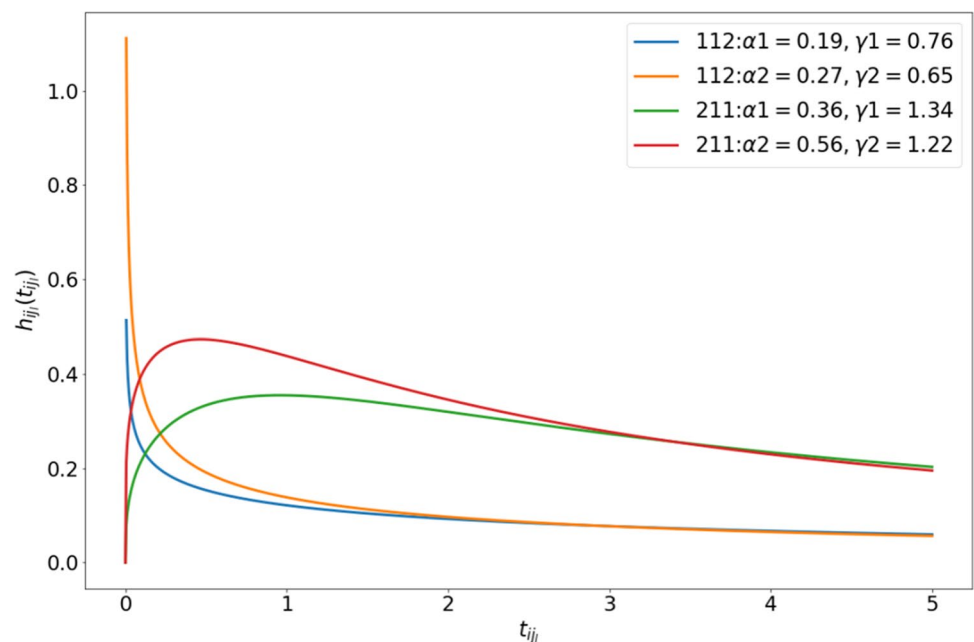
consumer groups. Specifically, the form of their browsing path characteristics in the related companies follows a single peak distribution. The hazard function value of the relevant company of these consumer groups is smallest at time 0, implying that when these groups of consumers enter the website, no matter which company they browse, they don't have a very clear goal to browse and compare companies at the beginning. In the first spell, when browsing duration increases, the distribution of hazard functions increases monotonously and reaches a turning point near the end of the first spell, and then the distribution of hazard functions gradually decreases. It shows that although a gradually increasing desire to browse and compare companies can be still observed, a turning point is reached near the end of the first spell, and then these consumer groups' desire to browse and compare to related companies gradually decreases. By contrast, for all the other types of consumers,

their hazard function value is the largest at time 0, but when the spell increases, the value of the hazard function gradually decreases. These results indicate that for all consumers other than the above-mentioned groups, a comprehensive comparison between companies starts immediately when they enter the website, while the comparisons fade down as time goes on.

4.3 Robustness Tests

4.3.1 Overall Consumers' Browsing Behavior

In line with prior studies (Bhatnagar et al., 2017; Chintagunta & Haldar, 1998), we test the robustness of our results by assuming that the spell data follow the Weibull distribution. Two distribution (Log-logistic and Weibull) models are used to analyze the influence of two factors, i.e.,

Table 8 Results of the overall consumers' browsing behavior robustness tests

Log-logistic	WAIC	Weibull	WAIC
Log-logistic-dependence for Category 1	56990.50	Weibull-*dependence* for Category 1	136125.00
Log-logistic-independence for Category 1	57850.20	Weibull-*independence* for Category 1	136303.40
Log-logistic-dependence for Category 2	42955.90	Weibull-*dependence* for Category 2	90339.30
Log-logistic-independence for Category 2	43448.00	Weibull-*independence* for Category 2	90615.98

Table 9 Consumer characteristics of browsing related companies in product category 1

Age	Gender	Membership level	No. of consumers (664)
≤ 35 (1)	*Male (0)*	*non-VIP (1)*	32
		VIP (2)	46
	Female (1)	*non-VIP (1)*	36
		VIP (2)	18
36–45 (2)	*Male (0)*	*non-VIP (1)*	85
		VIP (2)	141
	Female (1)	*non-VIP (1)*	54
		VIP (2)	49
≥ 46 (3)	*Male (0)*	*non-VIP (1)*	62
		VIP (2)	68
	Female (1)	*non-VIP (1)*	41
		VIP (2)	32

distribution form of spell data and dependence of browsing paths, on consumers' actual browsing. The robustness of the findings is tested according to the best-fit effect with the same data set for both product categories. In Table 8 we summarize our results by comparing the WAIC values of four different models in each product category. We

can see that incorporating the dependence of browsing paths performs better for explaining overall consumers' browsing behavior, and the spell data follow the Log-logistic distribution, which verifies the robustness of our findings.

4.3.2 Consumer Characteristics

To study the heterogeneity of consumers' browsing paths of related companies in the same product category, we classified all consumers according to the available demographic information including i) age (3 available types), ii) gender (2 available types), and iii) membership level (2 available levels). Accordingly, consumers can be divided into 12 (3 * 2 *2) types, and the number of consumers in each type is shown in Table 9.

Table 10 captures the browsing path characteristics of the 12 groups. Our analyses demonstrate that *Gamma* values have three types of distributions:

1. both *Gamma* values are greater than 1, referring to the only consumer group 301
2. only one of the *Gamma* values is larger than 1, referring to consumer groups 111 and 201
3. both *Gamma* values are less than 1, referring to all other consumer groups in Table 10.

Table 10 Consumers' browsing path characteristics of companies E & F

Parameter	101 Mean (s.d.)	102 Mean (s.d.)	111 Mean (s.d.)	112 Mean (s.d.)	201 Mean (s.d.)	202 Mean (s.d.)	211 Mean (s.d.)	212 Mean (s.d.)	301 Mean (s.d.)	302 Mean (s.d.)	311 Mean (s.d.)	312 Mean (s.d.)
Alpha1 (scale-company E)	0.21 (0.02)	0.23 (0.02)	0.37 (0.17)	0.18 (0.02)	0.36 (0.14)	0.24 (0.01)	0.19 (0.01)	0.18 (0.01)	0.59 (0.04)	0.20 (0.01)	0.21 (0.02)	0.21 (0.02)
Gamma1 (shape-company E)	0.70 (0.02)	0.73 (0.02)	1.09 (0.37)	0.74 (0.04)	1.14 (0.41)	0.75 (0.01)	0.71 (0.02)	0.72 (0.02)	1.61 (0.05)	0.73 (0.02)	0.69 (0.02)	0.75 (0.03)
Alpha2 (scale-company F)	0.35 (0.03)	0.30 (0.02)	0.43 (0.15)	0.27 (0.04)	0.41 (0.13)	0.28 (0.01)	0.28 (0.02)	0.30 (0.03)	0.86 (0.09)	0.35 (0.02)	0.34 (0.03)	0.22 (0.02)
Gamma2 (shape-company F)	0.43 (0.03)	0.55 (0.02)	0.93 (0.45)	0.56 (0.04)	0.85 (0.34)	0.55 (0.01)	0.53 (0.02)	0.46 (0.02)	1.38 (0.07)	0.47 (0.02)	0.46 (0.02)	0.57 (0.03)
Beta (Browsing-Path Dependence)	0.98 (0.02)	0.97 (0.03)	0.97 (0.03)	0.90 (0.09)	0.99 (0.01)	0.99 (0.01)	0.98 (0.02)	0.98 (0.02)	0.99 (0.01)	0.98 (0.02)	0.98 (0.02)	0.97 (0.03)

We compare the browsing behavior of heterogeneous groups according to the above distribution of *Gamma* values by incorporating the following three groups in the detailed analysis: 301 in type (1), 111 and 201 in type (2), and 202 as a random example in type (3). The Hazard Function of these three groups is displayed in Fig. 4.

The three-digit number in Fig. 4 refers to the focal consumer group. As each group browses two related companies at the same time, there are two curves of browsing path characteristics for each group: *Alpha1* and *Gamma1* represent the scale and shape parameters of company *E*, while *Alpha2* and *Gamma2* represent the scale and shape parameters of company *F* respectively. Several visual inferences can be drawn from Fig. 3. Group 301 is the only one having both *Gamma* values greater than 1, suggesting that non-VIP male consumers over 46 years of age are different from all other consumer groups. Specifically, the form of their browsing path characteristics in the related companies follows a single peak distribution. The hazard function value of the relevant company of this group is smallest at time 0, implying that when this group enters the website, no matter which company they browse, they do not have a very clear goal to browse and compare companies at the beginning. In the first spell, when browsing duration increases, the distribution of hazard functions increases monotonously and reaches a turning point near the end of the first spell, and then the distribution of hazard functions gradually decreases. It shows that although a gradually increasing desire to browse and compare companies can be still observed, a turning point is reached near the end of the first spell, and

then this group's desire to browse and compare to related companies gradually decreases.

Our findings indicate the single peak distribution also applies to the consumer groups labeled 111 and 201 as shown in Fig. 3. However, unlike the browsing behavior characteristics of the consumer group labeled 301, the hazard function value for the groups labeled 111 and 201 at time 0 is the largest for *F*, and the smallest for company *E*. When the spell increases, the risk function value of *E* is gradually greater than that of *F*. These results demonstrate when the groups labeled 111 and 201 enter the website, they are more interested in one certain company (i.e., company F), but as the browsing process goes on, it gradually develops a stronger interest in another company (i.e., company E), implying that these two consumer groups are easily interfered by external information. The hazard function value for all the other types of consumers is the largest at time 0, but when the spell increases, the value of the hazard function gradually decreases. These results indicate that for all consumer groups other than the above-mentioned groups labeled 301, 111, and 201, a comprehensive comparison between companies start immediately when they enter the website, while the comparisons fade down as time goes on.

5 Discussion

5.1 Key Findings

We use duration analysis to explore how path dependence can better explain consumers' subsequent behavior. We develop a multiple-spell competing hazard model that

Fig. 4 Hazard function of different consumer groups

Legend:
- 112: $\alpha1 = 0.18$, $\gamma1 = 0.74$
- 112: $\alpha2 = 0.27$, $\gamma2 = 0.56$
- 201: $\alpha1 = 0.36$, $\gamma1 = 1.14$
- 201: $\alpha2 = 0.41$, $\gamma2 = 0.85$
- 202: $\alpha1 = 0.24$, $\gamma1 = 0.75$
- 202: $\alpha2 = 0.28$, $\gamma2 = 0.55$
- 301: $\alpha1 = 0.59$, $\gamma1 = 1.61$
- 301: $\alpha2 = 0.86$, $\gamma2 = 1.38$

y-axis: $h_{iji}(t_{iji})$ x-axis: t_{iji}
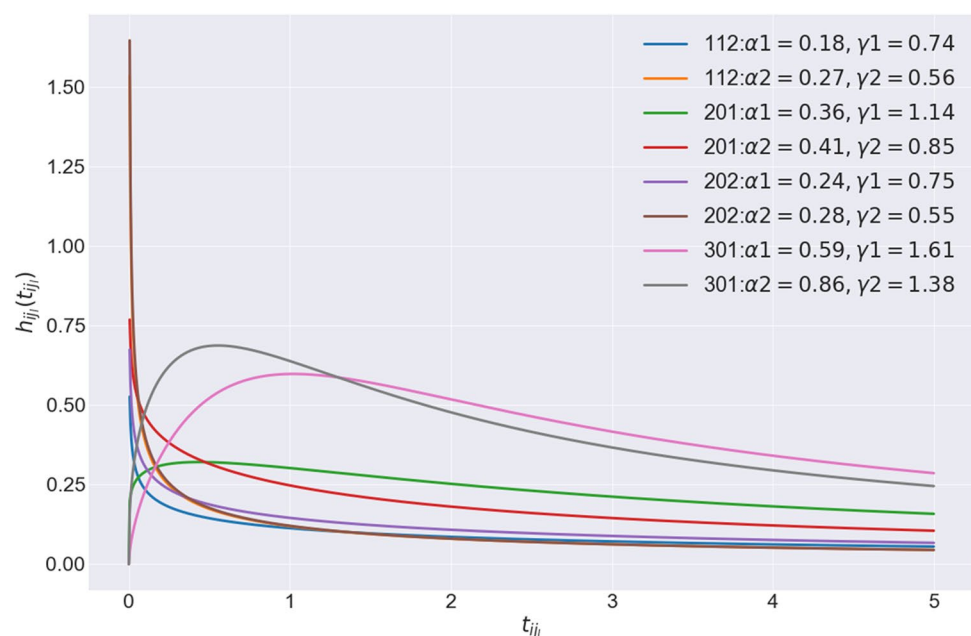
incorporates the dependence of consumers' browsing paths in related companies. Through the analysis of overall consumers' browsing behavior, we find that compared to the bivariate independent model, the multivariate model that considers the dependence of consumers' sequential browsing paths can accurately describe their actual behavior. Moreover, the Log-logistic distribution of spell data fits better with their actual behavior than the Weibull distribution. Particularly, the dependence on the browsing path of related companies is the highest on the first calendar day, and then it gradually decreases. Through the analysis of the heterogeneous browsing behavior of consumers, we find that when different groups browse related companies, their behavior patterns are quite different, and the degree of this difference is related to the company's position in the market. According to parameter estimates of heterogeneous consumer groups, for groups with unclear goals, a personalized recommendation service strategy should be implemented near the end of the first spell to increase the recommendation conversion rate. For other groups, as they have clear goals when they browse, online platforms should implement fewer personalized recommendations for these consumer groups to better utilize the resources.

5.2 Theoretical and Managerial Implications

Our findings contribute to research on online service response strategy, personalized recommendations, and duration analysis. Online service response strategy includes both in- and post-purchase components, however, prior studies have explored the topic only from a consumer review perspective after purchase. Compared to physical service, online one lacks face-to-face contact, which may result in greater perceived risks. Thus, timely online service response during purchasing is conducive to improving purchasing conversion rates. Contributing to this line of work, we incorporate a perspective of consumers' online browsing behavior. Although prior studies considered the browsing behavior of consumers (e.g., Bhatnagar et al., 2017), they have not considered the browsing-path dependence of related companies. Moreover, we explain the browsing behavior and the spell of the browsing behavior from the perspective of the heterogeneity of consumer groups.

For the field of personalized recommendation, current research usually studies how to improve the accuracy of personalized recommendation from the perspective of algorithms (Chang & Jung, 2017; Gan et al., 2019). One of the biggest challenges for recommendation algorithms is data sparsity and the cold start problem (Bunnell et al., 2020). To effectively solve this problem, we use information from heterogeneous consumer groups. In addition, equally important to the recommendation algorithm is the

timing of recommendations, since inappropriate recommendations are not conducive to purchase conversion. Thus, we first identify the types of heterogeneous consumer groups that are suitable for personalized recommendation and those that are not, and further analyze the timing of recommendation. Thus, we further enrich the theoretical underpinnings of personalized recommendations, while also presenting a methodological contribution by incorporating duration analysis, which is used to analyze consumers' purchasing behavior (Chintagunta & Haldar, 1998) but is less frequently adopted in information systems (IS) (Bhatnagar et al., 2017). Compared to the online browsing behavior of consumers, their purchasing behavior cannot reveal their real-time intentions for dynamically adapting the service response strategy of e-commerce platforms.

When it comes to the implications for managerial practice, our results provide a solid reference point for online marketplaces to optimize differentiated service response strategies. Specifically, our findings provide rich insights into consumer differentiation. For groups with a Gamma value of the hazard function greater than 1, as they do not have clear browsing goals, online marketplaces should focus on providing personalized recommendations near the end of the first spell to increase their conversion rates. For the other groups that have clear browsing goals, online marketplaces should implement fewer personalized recommendations to better utilize their computational resources.

5.3 Limitations and Future Research

Although we followed a structured research design, there are limitations that we need to acknowledge, and which present opportunities for future research. In line with the literature, as well as to simplify our data processing and analyses, we aggregated consumers' behavior to calendar days, which might result in coarser granularity in the subsequent analysis of consumer access time. To further address this issue, we suggest that future research should use various and multiple time windows such as minutes or hours for the data analysis. Additionally, limited by the model and calculation choices, we have only used two related companies to compare the fitting effect. In practice, however, consumers may compare more than two companies during their shopping journey. We, therefore, encourage future research to explore the topic with more combinations of products, companies, and categories. Furthermore, when analyzing consumer heterogeneity, we take the group as a unit. In practice, however, the behavior of individual consumers is heterogeneous. Therefore, we recommend that future research should use individual consumers as a unit when analyzing the heterogeneity of browsing path dependence. Ideally, future research should explore whether the personalized recommendation strategy

of heterogeneous consumer groups is beneficial to improve the revenue of a company, as well as the performance of companies in general in terms of other indicators. Finally, we adopted the product categorization of the focal e-commerce platform without further interference in order to enhance the practical implications of our findings. To further contribute to this line of work, future research could form categories based on the nature of products through, for instance, the product involvement concept (e.g., Drossos et al., 2013) or by distinguishing between hedonic and utilitarian ones (Lim & Ang, 2008). The novel avenues for future research that stem from our work can be extended beyond the context of e-commerce platforms. For instance, online recommendations can be beneficial for e-government websites (e.g., Angelopoulos et al., 2010; Kitsios et al., 2009) and can further be applied to policymaking (e.g., Georgiadou et al., 2020). Moreover, our work could also be extended through research on the use of intimate personal data by e-commerce platforms (e.g., Angelopoulos et al., 2021), and has extensions related to cybersecurity threats (e.g., Janse et al., 2017; Ou et al., 2022). We, therefore, encourage future research to further explore these topical and timely areas, in line with the extant IS research agenda (Struijk et al., 2022).

6 Conclusion

In this paper, we examined how the sequential browsing behavior of consumers can enable targeted marketing strategies for e-commerce platforms, by using clickstream data. We deploy duration analysis to explore how path dependence can better explain consumers' sequential browsing behavior in various product categories and characterize the sequential browsing behavior of heterogeneous consumer groups. In doing so, we demonstrate that sequential browsing path dependence can explain with high accuracy the behavior of consumer. Our findings provide nuanced expositions for strategic decision-making on e-commerce platforms and open up avenues for future research.

Declarations

References

Angelopoulos, S., Kitsios, F., & Papadopoulos, T. (2010). New service development in e-government: Identifying critical success factors. *Transforming Government: People, Process and Policy, 4*(1), 95–118.

Angelopoulos, S., Brown, M., McAuley, D., Merali, Y., Mortier, R., & Price, D. (2021). Stewardship of personal data on social networking sites. *International Journal of Information Management, 56*, 102208.

Balan, U. M., & Mathew, S. K. (2021). Personalize, summarize or let them read? A study on online word of mouth strategies and consumer decision process. *Information Systems Frontiers, 23*(3), 627–647.

Behera, R. K., Bala, P. K., & Ray, A. (2021). Cognitive Chatbot for personalised contextual customer service: Behind the scene and beyond the hype. *Information Systems Frontiers*, 1-21. https://doi.org/10.1007/s10796-021-10168-y

Bhatnagar, A., Sen, A., & Sinha, A. P. (2017). Providing a window of opportunity for converting eStore visitors. *Information Systems Research, 28*(1), 22–32.

Bronnenberg, B. J., Kim, J. B., & Mela, C. F. (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing Science, 35*(5), 693–712.

Bucklin, R. E., & Sismeiro, C. (2009). Click here for internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing, 23*(1), 35–48.

Bunnell, L., Osei-Bryson, K. M., & Yoon, V. Y. (2020). RecSys issues ontology: A knowledge classification of issues for recommender systems researchers. *Information Systems Frontiers, 22*(6), 1377–1418.

Chang, W. L., & Jung, C. F. (2017). A hybrid approach for personalized service staff recommendation. *Information Systems Frontiers, 19*(1), 149–163.

Chen, Y., & Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science, 63*(12), 4345–4365.

Chintagunta, P. K., & Haldar, S. (1998). Investigating purchase timing behavior in two related product categories. *Journal of Marketing Research, 35*(1), 43–53.

Chu, V. W., Wong, R. K., Chi, C. H., Zhou, W., & Ho, I. (2017). The design of a cloud-based tracker platform based on system-of-systems service architecture. *Information Systems Frontiers, 19*(6), 1283–1299.

Drossos, D. A., Giaglis, G. M., Vlachos, P. A., Zamani, E. D., & Lekakos, G. (2013). Consumer responses to SMS advertising: Antecedents and consequences. *International Journal of Electronic Commerce, 18*(1), 105–136.

Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika, 47*(3/4), 307–323.

Gan, M., Sun, L., & Jiang, R. (2019). GLORY: Exploration and integration of global and local correlations to improve personalized online social recommendations. *Information Systems Frontiers, 21*(4), 925–939.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Georgiadou, E., Angelopoulos, S., & Drake, H. (2020). Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes. *International Journal of Information Management, 51*, 102048.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*(1), 1593–1623.

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science, 28*(2), 320–335.

Janse, N., Ou, C. X., Angelopoulos, S., Davison, R. M., & Jia, J. W. (2017). Do security breaches matter to consumers?. *In the proceedings of the 17th International Conference on Electronic Business*, 321.

Johnson, N. L., & Kott, S. (1975). On some generalized Farlie-Gumbel-Morgenstern distributions. *Communications in Statistics-Theory and Methods, 4*(5), 415–427.

Johnson, E. J., Bellman, S., & Lohse, G. L. (2003). Cognitive lock-in and the power law of practice. *Journal of Marketing, 67*(2), 62–75.

Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). On the depth and dynamics of online search behavior. *Management Science, 50*(3), 299–308.

Karimi, S. (2021). Cross-visiting behaviour of online consumers across retailers' and comparison sites, a macro-study. *Information Systems Frontiers, 23*, 531–542.

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2011). Mapping online consumer search. *Journal of Marketing Research, 48*(1), 13–27.

Kitsios, F., Angelopoulos, S., & Zannetopoulos, J. (2009). Innovation and e-government: an in depth overview on e-services. In *Handbook of Research on Heterogeneous Next Generation Networking: Innovations and Platforms* (pp. 415–426). IGI Global.

Lim, E. A. C., & Ang, S. H. (2008). Hedonic vs. utilitarian consumption: A cross-cultural perspective based on cultural conditioning. *Journal of Business Research, 61*(3), 225–232.

Manouchehri, N., Nguyen, H., Koochemeshkian, P., Bouguila, N., & Fan, W. (2020). Online Variational learning of Dirichlet process mixtures of scaled Dirichlet distributions. *Information Systems Frontiers, 22*(5), 1085–1093.

Moe, W. W., & Fader, P. S. (2004a). Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing, 18*(1), 5–19.

Moe, W. W., & Fader, P. S. (2004b). Dynamic conversion behavior at E-Commerce sites. *Management Science, 50*(3), 326–335.

Montgomery, A. L. (2001). Applying quantitative marketing techniques to the internet. *Interfaces, 31*(2), 90–108.

Ou, C. X., Zhang, X., Angelopoulos, S., Davison, R. M., & Janse, N. (2022). Security breaches and organization response strategy: Exploring consumers' threat and coping appraisals. *International Journal of Information Management, 65*, 102498.

Park, C. H. (2017). Online purchase paths and conversion dynamics across multiple websites. *Journal of Retailing, 93*(3), 253–365.

Park, Y.-H., & Fader, P. S. (2004). Modeling browsing behavior at multiple websites. *Marketing Science, 23*(3), 280–303.

Proserpio, D., & Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science, 36*(5), 645–665.

Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an E-Commerce Web Site: A task-completion approach. *Journal of Marketing Research, 41*(3), 306–323.

Struijk, M., Ou, C. X. J., Davison, R. M., & Angelopoulos, S. (2022). Putting the IS back into IS research. *Information Systems Journal, 32*(3), 469–472. https://doi.org/10.1111/isj.12368

Su, Q., & Chen, L. (2015). A method for discovering clusters of E-Commerce interest patterns using click-stream data. *Electronic Commerce Research and Applications, 14*(1), 1–13.

Udo, G. J., Bagchi, K. K., & Kirs, P. J. (2010). An assessment of customers' e-service quality perception, satisfaction and intention. *International Journal of Information Management, 30*(6), 481–492.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using Leave-One-out Cross-Validation and Waic. *Statistics and Computing, 27*(5), 1413–1432.

Wang, Y., Lo, H. P., & Yang, Y. (2004). An integrated framework for service quality, customer value, satisfaction: Evidence from China's telecommunication industry. *Information Systems Frontiers, 6*(4), 325–340.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*, 3571–3594.

Zuo, M., Liu, H., Zhu, H., & Gao, H. (2019). Dynamic property of consumer-based brand competitiveness (CBBC) in human interaction behavior. *Industrial Management & Data Systems, 119*(6), 1223–1241.

Holloway, B. B., & Beatty, S. E. (2008). Satisfiers and dissatisfiers in the online environment: A critical incident assessment. *Journal of Service Research, 10*(4), 347–364.

Meihua Zuo is a lecturer in Huizhou University, China. She received her PhD from the School of Management from Guangdong University of Technology, China. Her research interests include consumer behavior analysis, business intelligence, social network, and brand competition of data mining. Her work has been published in journals like Industrial Management & Data Systems (IDMS), and Journal of Control & Decision, and presented at conferences like International Conference on Information Systems (ICIS), the Academy of Management (AOM), and the International Conference on Electronic Commerce (ICEC). She has also participated in several Natural Science Foundation of China (NSFC) projects.

Spyros Angelopoulos is an Associate Professor at Durham University Business School. Previously, he worked at Tilburg University, the University of Lugano, and the University of Nottingham. He holds a PhD from Warwick Business School and has a background in Engineering and Management. His research focuses on users' behavior on online platforms, organizational adaptation during digital transformation, and on security and privacy issues of users on online platforms. His publications include book chapters, journal papers, and conference presentations. He is Guest Editor in Special Issues at the Journal of Operations Management and the International Journal in Information Management. He was selected as 'Faculty Expert' by Google, for his commitment to leading transformation in Cloud Computing research and education.

Zhouyang Liang is an experimenter at the Guangdong University of Technology, China, where he also received his PhD from. His research interests include consumer behavior analysis, electronic commerce, Bayesian inference, and recommendation system. His work has been published in journal like Journal of Control and Decision. As well as, and conferences like International Conference on Information Systems (ICIS), Annual Meeting of the Academy of Management (AOM), and International Conference on Electronic Commerce (ICEC). He has participated in several projects of the Natural Science Foundation of China (NSFC).

Carol X. J. Ou is a Professor of Digital Transformation and Information Management at the Management Department in Tilburg University, the Netherlands. She is currently serving as the Head of Department. Her research interests include digital transformation, applied business intelligence, computer-mediated communication, social commerce, smart recommendation agents and knowledge management. Her publications have appeared in journals such as Communications of the ACM, Decision Support Systems, Information & Management, Information Technology & People, Journal of AIS, Journal of the American Society for Information Science and Technology, International Journal of Human–Computer Studies, Information Systems Journal, Journal of IT Management, and MIS Quarterly. She serves as a senior editor for Information Systems Journals, Information & Management, and IT & People. She has chaired tracks in major Information Systems conferences such as ICIS, ECIS and PACIS. She is also a Certified IS Auditor and an Academic Advocate of IS Audit and Control Association. Over the past years, she has led or participated in project grants of over €1.8 million.