

Will this search end up with booking? Modeling airline booking conversion of anonymous visitors

Misuk Lee

*Albers School of Business and Economics, Seattle University,
Seattle, Washington, USA*

Booking
conversion of
anonymous
visitors

237

Received 14 November 2019
Revised 1 April 2020
7 May 2020
Accepted 28 May 2020

Abstract

Purpose – Over the past two decades, online booking has become a predominant distribution channel of tourism products. As online sales have become more important, understanding booking conversion behavior remains a critical topic in the tourism industry. The purpose of this study is to model airline search and booking activities of anonymous visitors.

Design/methodology/approach – This study proposes a stochastic approach to explicitly model dynamics of airline customers' search, revisit and booking activities. A Markov chain model simultaneously captures transition probabilities and the timing of search, revisit and booking decisions. The suggested model is demonstrated on clickstream data from an airline booking website.

Findings – Empirical results show that low prices (captured as discount rates) lead to not only booking propensities but also overall stickiness to a website, increasing search and revisit probabilities. From the decision timing of search and revisit activities, the author observes customers' learning effect on browsing time and heterogeneous intentions of website visits.

Originality/value – This study presents both theoretical and managerial implications of online search and booking behavior for airline and tourism marketing. The dynamic Markov chain model provides a systematic framework to predict online search, revisit and booking conversion and the time of the online activities.

Keywords E-commerce, Markov chain model, Airline booking, Clickstream data, Airline booking, Search to booking conversion, Online behavior

Paper type Research paper

1. Introduction

Over the past two decades, online booking in the tourism industry has grown exponentially and become a predominant distribution channel of tourism products.

This trend is expected to continue, and according to Phocuswright (2017), by 2020, online channels (including mobile) will account for 49% of the total US travel booking. As online sales have become more important in the tourism industry, it has become more critical to understand booking conversion behavior to capture more revenue opportunities. How to promote customer engagement and conversion on the path to booking emerges as the most important task in tourism marketing.

In the tourism industry, online promotions are common, but the costs of promotions are relatively high. Predicting latent booking conversion is particularly important for tourism



marketing. In the competitive environment, failure to understand latent demand may result in lost revenue opportunities.

While there have been huge interests in analyzing online booking for tourism products, there are few published studies on online browsing behavior of anonymous visitors. With little information on customer profiles or demographics, predicting booking conversion of anonymous visitors is a very challenging task.

The objective of this study is to develop a stochastic model to comprehend many complexities of anonymous visitors' online booking behavior. We propose a framework to explore not only purchase conversion but also entire search and visit processes. We demonstrate our theoretical model using clickstream data obtained from a low-cost carrier. Through the empirical study, we investigate key research questions on anonymous customers' online behavior.

We explore within-site browsing activities of anonymous airline customers using clickstream data, and model dynamic conversion behavior across search, revisit and booking decisions. Moreover, along with these three decisions estimated by conversion probabilities, we consider timing of the decisions. To simultaneously capture customer decisions and intra-/inter-visit durations, we propose a stochastic model that integrates discrete choice theory with a Markov chain.

Our stochastic model provides a framework to explore key research questions on booking conversion behavior. First, we investigate price effects on willingness-to-purchase as well as customer engagement within a website, captured as search and revisit probabilities. While one can easily imagine that low prices may increase customers' booking propensities, there have been few (if any) studies that investigated the price effects on customers' search and revisit decisions. Second, we examine how customers build website engagement (or website stickiness), and how search activities and time spent on the website may affect the search, booking and revisit decisions. While these two research questions concern the three decisions (search, booking and revisit) in online browsing, our stochastic model allows us to examine decision timing revealed in page view time and inter-visit duration. The third research question is how much customers spend in viewing a Web page and if they change their behavior in terms of page view time. Lastly, we investigate inter-arrival time between visits and customers' revisit intention through the inter-visit time.

Using the Markov chain model, we examine how our model can help answer the above research questions and understand dynamics of anonymous visitors' online browsing behavior. We demonstrate our framework by applying the Markov chain model to clickstream data from an airline booking website.

The rest of the paper is organized as follows. First, we present a literature review to understand previous research streams on online conversion behaviors. Next, we describe airline booking clickstream data used in this study. Then, we present our Markov chain model for booking conversion, followed by empirical results obtained from the clickstream data. We conclude by describing discussions and future research.

2. Literature review

2.1 Online browsing behavior

2.1.1 Search to purchase conversion. Our work relates to various streams of existing literature on online browsing behaviors and purchase conversion. Many studies including Chatterjee *et al.* (2003), Moe and Fader (2004), Iwanaga *et al.* (2016), Park and Park (2016) and Mokryn *et al.* (2019) investigated online search activities and their effects on purchase conversions. Chatterjee *et al.* (2003) developed a model to capture search and conversion behaviors in online banner advertising. Moe and Fader (2004) proposed an individual-level

statistical model to understand evolving visiting behavior. Moreover, they examined the relationship between visit frequency and purchase probability. [Park and Park \(2016\)](#) suggested a model to predict online store visit and purchase probabilities. They specifically investigated customer visit patterns and their impacts on purchase dynamics across store visits. [Iwanaga et al. \(2016\)](#) studied the relationship between customers' page views and the product choice-probabilities on an online shopping site. [Mokryn et al. \(2019\)](#) suggested the use of products' popularity trends and visit's temporal information to predict the purchase intention of anonymous visitors. [Yeo et al. \(2020\)](#) specifically investigated revisit customers' purchase conversion. They developed product-level and customer-level models for conversion predictions and predictability. While these studies have focused on understanding purchase conversion, our approach adopts more comprehensive framework to incorporate search, revisit and purchase decisions across entire online shopping process into timing of such decisions.

2.1.2 Online shoppers' browsing intension. In terms of understanding customers' intentions of purchasing, several previous studies have noticed different objectives of online shoppers. [Novak et al. \(2003\)](#) proposed two distinct consumption behaviors in e-commerce. They found an evident of different online flow experiences for goal-directed behavior from experiential consumption behavior. [Moe \(2003\)](#) focused on two different types of online shoppers. They found that some shoppers are focused in looking for a specific product while some other shoppers show window shopping patterns. [Su and Chen \(2015\)](#) demonstrated that user's browsing behaviors such as browsing paths, the frequency of page visits and the time spent on each category, represent a comprehensive reflection of their interests. [Ko \(2020\)](#) also explored the influence paths driving experiential browsing and goal-directed shopping intentions in the e-commerce context.

Capturing consumer browsing intension may have a very important managerial implication in the context of real-time promotion and customized marketing. As [Novak et al. \(2003\)](#) suggested, we hypothesize that there may exist two types of online shopping intentions – goal-oriented and experiential. We hypothesize that these two online browsing intentions may be captured through interval between visits. Customers who come back to the website after relative short duration may be more goal oriented. On other hand, customers who revisit after long interval may be more likely to be experiential.

2.1.3 Timing of search and visit. Another stream of research related to our work is on investigations of page view durations and visit durations. [Bucklin and Sismeiro \(2003\)](#) and [Johnson et al. \(2004\)](#) have examined whether page view and visit durations decrease as customers gain more knowledge on the website. [Bucklin and Sismeiro \(2003\)](#) developed a model of website browsing behavior and found that as customers visited repeatedly, they tended to request fewer page views and reduce total visit durations. [Johnson et al. \(2004\)](#) observed that total visit duration decreased over time. However, their study did not explicitly consider intra-visit search activities such as page view durations and search depth. [Bhatnagar et al. \(2017\)](#) explored dynamics of website visits in the context of advertising tool – banner ad or search engine. They found that inter-visit duration affects search behavior in the subsequent visit. Their empirical results showed that as the interval between visits increases, purchase probability tends to decrease.

Our research investigates duration of search and interval between searches within and across visits, through which we can have more comprehensive view on dynamics of search behavior. We explore how online shoppers change page view time and visit duration as [Bucklin and Sismeiro \(2003\)](#) and [Johnson et al. \(2004\)](#) suggested. Moreover, our model integrates effects of search and visit duration on purchase, revisit and search decisions.

2.2 Online booking behavior for tourism products

As online has become a predominant distribution channel of tourism products, there have been growing interests on analyzing online purchasing in the tourism industry. However, as [Morosan and Bowen \(2017\)](#) pointed out, studies on online purchase behavior have a strong orientation toward self-reported survey data or controlled experiment data. There are only a few published studies on analyzing online booking transactions or clickstream data for tourism products. [Cezar and Ögüt \(2016\)](#) explicitly analyze conversion rate in online hotel booking. They investigated impacts of review rating, recommendation and search listings on booking conversion. Moreover, they found that room price and hotel size are negatively associated with conversion rate. [Chen and Yao \(2017\)](#) proposed a model for consumers' decisions of search and refinement such as filtering and sorting. Their empirical results obtained from click-stream data of online hotel bookings showed that refinement tools encourage more searches and enhance the utility of purchased products significantly. [Xie and Lee \(2019\)](#) investigated how informational cues such as discount price, promotions, brand and quality rating influence search, click-through and booking conversion in an online hotel search process. Using click-stream data from Expedia, this study found that consumers are more likely to click through hotels with higher ratings and to book hotels with consumer generated ratings. Moreover, consumers tend to highly engage in discounted price and promotion. [Zhang et al. \(2019\)](#) investigated a trade-off between immediate booking and continuing information search with delayed purchase decision. Based on data set of online peer reviews and consumer reservation records, they investigated the effects of online peer reviews on consumers' timing of booking a restaurant. They found that rating valence, rating variation and review content richness tend to promote potential consumers to book a restaurant earlier.

While many studies have explored customers' browsing behavior for online shopping, they investigated individual aspects of online behaviors. This research provides comprehensive framework for understanding many complexities of online shopping behavior – search and visit timing, browsing intention and purchase conversion.

Investigating consumer search behaviors has especially important managerial implications in the context of predicting purchase conversion. While price, often measured by discount rate, may provide presumably key motivation for purchase decision, very few studies have addressed price effects on online browsing patterns. We explore how price (discount) is correlated with dynamics of online search behavior throughout search, revisit and purchase.

3. Data

Recent studies on individual customer's online behavior are primarily because of rich data sets from website browsing logs – clickstream data. Clickstream data records a user's actions on a specific website, and thus it is often called digital footprints. Therefore, it can show in detail what a user does, which pages he/she goes and when/in what order he/she takes an action.

To describe clickstream data, we define three terms structured in a hierarchical order. A *search* is defined as a customer click on a page and it is the lowest level of activity in this research. A *visit* is a series of searches requested within a specific period. While search and visit are the fundamental elements of online activities, a booking may be made after many searches, and often visits, spanning hours, days or even weeks. In fact, many of previous studies including [Chatterjee et al. \(2003\)](#), [Moe and Fader \(2004\)](#) and [Park and Park \(2016\)](#) investigated online behavior across visits. Lastly, a *user* is identified through a cookie, which is stored in a user's computer. Even though there exist some limitations, tracing cookies can be the most viable method to identify a user ([Trusov et al., 2016](#)). While most

users search for a single product, some search for multiple products, in which case the same user (cookie ID) is associated with multiple booking cycles.

To demonstrate our model, we have collected clickstream data from an online booking website for a low-cost carrier serving Australia and New Zealand. The data was collected over a two-year period from customers who visited webpages offering specific promotional products targeting primarily leisure travelers. By focusing on a specific type of leisure products, we may observe customers' browsing activities in a more controlled environment. Moreover, because of the characteristics of promotional products, price variations are large enough to evaluate price (discount) effects on search and purchase behavior. The data set used in this study contains 42,554 searches across 20,354 visits from 12,588 unique cookies (users). The number of bookings is 474, which yields 3.76% conversion rate per user.

4. Model

At a fundamental level, the objective of this research is to model dynamics of online booking activities within a website. More specifically, we model dynamic transitions among three decisions in online booking – search, revisit and booking. Figure 1 illustrates the three decisions as transitions among four navigation states – *search*, *booking*, *leave* and *no booking*. Once a customer searches for a product, he/she enters the *search* state. While in the *search* state, the customer decides if he/she keeps searching or not. If he/she requests additional page, he/she stays in the *search* state. If not, the customer may purchase or leave the website, entering the *booking* state or the *leave* state. Those customers who are in the *leave* state may come back to the website, which corresponds to revisit decision. Note that the *search* and the *leave* states are transient states, and the *booking* and *no booking* states are recurrent states. That is, a Markov chain for the search, revisit and booking activities will end up with either a *booking* or *no booking* state.

4.1 Search, revisit and booking probabilities

We model customer search to booking transitions (illustrated in Figure 1) as a Markov chain with the state space $S = \{search, booking, leave, no\ booking\}$. The transition from state i to j in a Markov chain is assumed to be driven by latent customer utility U_{ij} . The customer utility is further decomposed into an observed (exploratory) variable vector X_{ij} along with its coefficients Γ_i , and a random component ϵ_{ij} .

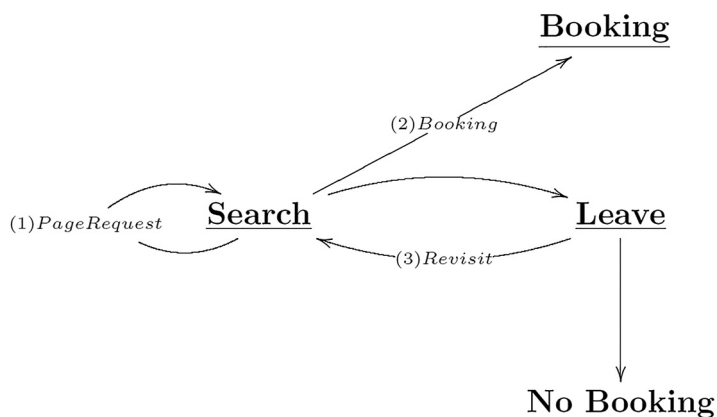


Figure 1. Transitions of search, revisit and booking activities

$$U_{ij} = \Gamma_i X_{ij} + \varepsilon_{ij}. \tag{1}$$

If ε_{ij} 's are assumed to be independent and identically distributed, and follow a standard Gumbel distribution, the conversion probability P_{ij} from state i to j follows a logistic distribution.

$$P_{ij} = \frac{e^{\Gamma_i X_{ij}}}{\sum_{m \in A_i} e^{\Gamma_i X_{im}}} \tag{2}$$

In our search model, the exploratory variable vector X_{ij} includes the price of a product and variables associated with search activities such as number of searches, number of visits, page view time and inter-visit duration. By estimating coefficient Γ_i on X_{ij} , we can understand effects of price and search activities on customer search, booking and revisit decisions.

We may interpret the above transition probabilities in the context of customers' search and booking decisions. As shown in [Figure 1](#), transitions between the navigation states correspond to a series of three binary decisions: whether to continue searching, whether to purchase before leaving the site and whether to revisit the site once having left without booking. Let us define choice probabilities of these three decisions:

- (1) f : probability of finishing a search;
- (2) b : probability of booking given that a customer ends searches; and
- (3) r : probability of revisit.

These three probabilities can be defined using the transition probabilities of a Markov chain derived in [equation \(2\)](#):

$$\begin{aligned} P_{SS} &= 1 - f \\ P_{SB} &= fb \\ P_{SL} &= f(1 - b) \\ P_{LS} &= r \\ P_{LN} &= 1 - r \end{aligned} \tag{3}$$

where S, B, L and N represent *search, booking, leave* and *no booking* states, respectively.

For example, the probability of keeping search, $1 - f$, corresponds to the transition within the *search* state (P_{SS}). The transition probability from *search* state to *booking* state (P_{SB}) can be decomposed into two probabilities. Once a customer decides to end search with probability f , he/she will purchase with probability b . Then, the unconditional booking probability (P_{SB}) is fb .

4.2 Decision time – page view and inter-visit duration

A continuous time Markov chain model allows us to estimate inter-arrival times between states through exponential distributions. In the context of online search activities, inter-arrival times are understood as page-view time and inter-visit time.

Let us denote ν the time between *search* states, which corresponds to page view time. Inter-arrival times of a continuous Markov chain assume to follow exponential distributions. Then, the probability of spending more than t in viewing a page is defined as:

$$P\{\nu > t\} = e^{-\lambda t}. \tag{4}$$

whereas λ is the rate between two consecutive page requests. [Figure 2](#) that presents the histogram of page-view times (ν) confirms that the frequency of ν decreases exponentially.

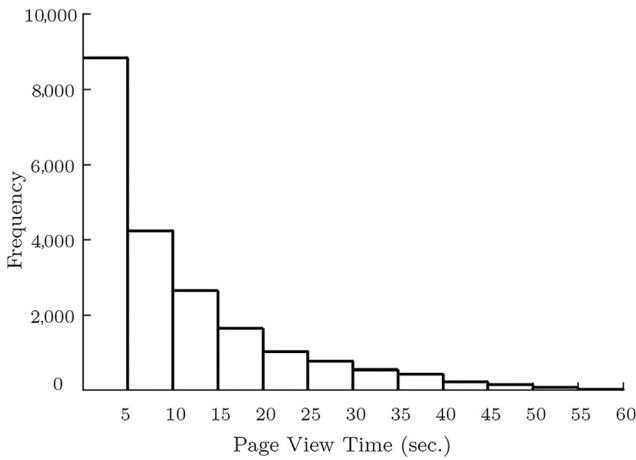


Figure 2. Histogram of page-view time

In modeling customers' behavior with respect to page view time, we hypothesize three different models. First, we assume that customers' page view time may be independent of search depth and number of visits, in which case λ is constant. Next, customers may dynamically change their page view durations as they search more and view more pages. If there are self-imposed time constraints on website usage, we may expect the page view time to decrease with search depth. Also, as customers accumulate more knowledge on the website, they may need less time in viewing pages (learning effect). If the time constraints and/or learning effects appear, we may assume λ to be a function of number of page requests (n), i.e. λ_n . Lastly, if the learning effects and/or self-imposed time constraints have impacts across visits, we may allow the inter-arrival rate λ_{vn} to change with search depth n and number of visits v .

An inter-arrival time between visits, τ , is time from the *leave* state to the *search* state. As in the page view time, τ is assumed to be exponentially distributed with rate μ .

$$P\{\tau < t|\mu\} = 1 - e^{-\mu t} \tag{5}$$

The histogram of revisit time τ in Figure 3 shows that τ decreases exponentially over page view time. Note that the x -axis (inter-visit time) in Figure 3 is not uniformly scaled.

Another insight we obtain from Figure 3 is that customer behavior in terms of revisit time is very diverse. Note that the time between visits in Figure 3 ranges from minutes to a year. To model the large variation of the revisit time, we model μ to be a random variable. More specifically we assume that μ follows a Gamma distribution of shape parameter θ and scale parameter δ . Then, the distribution of τ is derived as:

$$\begin{aligned} P\{\tau < t\} &= \int_{-\infty}^{\infty} (1 - e^{-\mu t}) dF(\mu) \\ &= 1 - (t/(t + \theta))^{-\delta}. \end{aligned} \tag{6}$$

On the other hand, previous studies such as Chatterjee *et al.* (2003) and Moe (2003) observed that there may exist two different motivations of online browsing. Goal-oriented customers

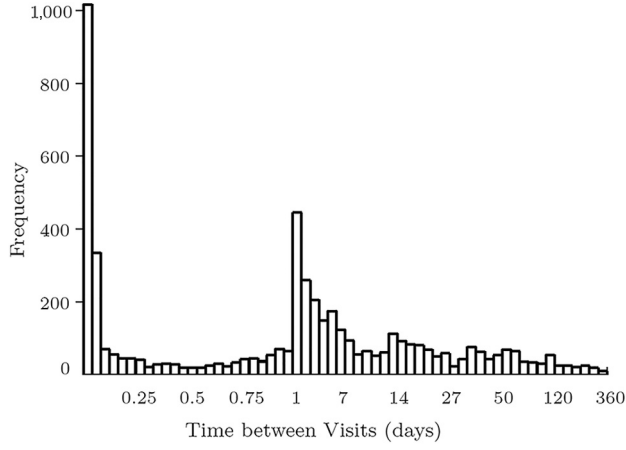


Figure 3.
Histogram of
revisit time

are more motivated to purchase and engage in a website. Those who visit the website without direct intention to purchase may be more experiential. If this theory is applied to our model, we may observe that goal-oriented customers may revisit after short time period while experiential customers need longer inter-visit durations to explore more options.

To investigate these distinct revisit behaviors, we hypothesize two populations of revisit time: those with short revisit time and those with long revisit time. With a certain probability α , a customer comes from the short inter-visit population, whose inter-arrival rate follows a Gamma distribution with parameters θ_s and δ_s . With probability $(1 - \alpha)$, a customer is from the long inter-visit population, whose inter-arrival rate follows a Gamma distribution with parameters θ_l and δ_l . That is, μ (inter-visit rate) may follow either $Gamma(\theta_s, \delta_s)$ with probability α or $Gamma(\theta_l, \delta_l)$ with probability $(1 - \alpha)$.

$$\mu \sim \begin{cases} Gamma(\theta_s, \delta_s) & \text{w.p. } \alpha_s \\ Gamma(\theta_l, \delta_l) & \text{w.p. } 1 - \alpha_s \end{cases} \quad (7)$$

Then, the distribution of τ is derived as:

$$P\{\tau < t\} = \alpha_s \left(1 - (t/(t + \theta_s))^{-\delta_s}\right) + (1 - \alpha_s) \left(1 - (t/(t + \theta_l))^{-\delta_l}\right). \quad (8)$$

5. Results

The continuous-time Markov chain model proposed in the previous section is applied to the airline booking clickstream data. Using the framework of Markov chains, we estimate customer search, visit and booking decision probabilities and inter-arrival times between pages.

5.1 Search, revisit and booking probabilities

The dynamic Markov chain model captures three decisions in online browsing – exit, revisit and booking. [Table 1](#) reports estimation results obtained from the clickstream data and

Table 1.

Estimation results:
exit, booking and
revisit probabilities

Variable name	Variable	Exit, f	Booking, b	Revisit, r
Discount rate (price effect)	d	-0.375 (0.000)	0.935 (0.000)	0.835 (0.000)
Page requests	n	-2.403 (0.000)	0.168 (0.020)	0.137 (0.000)
Page view time	ν	-2598.882 (0.000)		
	ω	8214.663 (0.000)		
Repeated visits	$1_{\{v = 1\}}$	-0.756 (0.000)	1.715 (0.000)	1.241 (0.000)
	$1_{\{v = 2\}}$	-0.830 (0.000)	2.337 (0.000)	2.122 (0.000)
	$1_{\{v = 3\}}$	-0.901 (0.000)	2.093 (0.000)	2.153 (0.000)
	$1_{\{v > 3\}}$	-0.789 (0.000)	2.039 (0.000)	3.252 (0.000)
Inter-visit time	τ	0.003 (0.000)	-0.0148 (0.001)	-0.002 (0.050)
Constant		0.893 (0.000)	-5.155 (0.000)	-2.338 (0.000)
	Observations	29836	14225	13905
	Log likelihood	-17070.915	-1353.218	-6303.1642

provide insights into price and search effects on customer online behavior revealed through the three decisions.

First of all, price effects on online activities seem very apparent. Table 1 shows that low prices (higher discount rates) increase not only direct booking propensities but also search and revisit probabilities. This implies that low prices promote consumer engagement throughout the entire search, revisit and booking decisions.

In explaining search effects on customer decisions, let us categorize the search effects into intra- and inter-visit effects. Intra-visit search effects are examined through number of searches and page-view time. Inter-visit search effects are investigated through number of visits and inter-visit durations.

Table 1 reports negative coefficients of search depth n and page-view time ν on f (probability of ending search). This implies that consumers become even more engaged in the website as they deepen their search and spend more time in viewing pages. Previous studies including Zauberman (2003) and Iwanaga *et al.* (2016) also suggested positive intra-visit search effects on customer engagement in online browsing. Moreover, the number of searches n has positive impacts on booking and revisit probabilities, which further supports positive intra-visit effects on customer engagement and booking.

In estimating f , we control for cumulative time spent in the current visit (ω). The positive coefficient on ω implies that the propensity to end searching increases with total time spent within a visit. Under the self-imposed time constraints, within-site stickiness behavior is observed within a visit.

The direction and significance of inter-visit search effects can be estimated by the coefficients of repeated visits and inter-visit durations. Table 1 shows that repeated visits are associated with higher search, booking and revisit probabilities. This positive inter-visit search effect is consistent with results from Bucklin and Sismeiro (2003) and Johnson *et al.* (2004). Table 1 reports negative coefficients of inter-visit time τ on booking probability b and negative coefficients on exit probability f . That is, longer inter-visit time may indicate that the customer is more likely to be goal-oriented and thus less likely to purchase and engage in the website.

5.2 Decision time – page view and inter-visit duration

Using the continuous time Markov chain model, we estimate exponential rates for page-view duration and inter-visit duration. Through the comparison of three different models for λ

(rates for page-view time), we may see if customers change their page view time dynamically as they navigate the website and accumulate information on products. More specifically, λ may be a function of the number of searches (n) and the number of visits (v). Table 2 presents Akaike information criterion (AIC) values for three models of λ . AIC is a measure of quality of statistical models and often used as a means of model selection. As AIC is computed as the number of parameters to be estimated minus goodness of fit (likelihood value of a given model), a model of lower AIC score is generally considered as a better model. In Table 2, we see a significant improvement in AIC value when we allow the page view time to change by the number of pages (Model of λ_n). Figure 2 further supports our hypothesis on different page view rates by the number of searches. That is, as customers search more, they tend to spend less time in viewing webpages. This phenomenon may be an evidence of learning effects or self-imposed time constraints for website navigation (Figure 4).

While developing models of inter-visit rates (μ), we hypothesized that inter-visit durations are very diverse as customers have distinct intentions for revisit. Some customers, who have more immediate needs for revisit may come back after short durations. Some may need longer inter-visit durations to explore more options. Table 3 reports AIC values of three different models for estimating μ . Significant improvement in AIC from the constant μ model to random μ model suggests that customers' intentions for revisit are indeed varied. Moreover, the model of two heterogeneous revisit rates achieves the best AIC score, which

Table 2.
Model fit: page-view time

Model	AIC
λ	32,631
$\lambda = \lambda_n$	30,481
$\lambda = \lambda_{vn}$	30,479

Figure 4.
Cumulative distribution of page view time by number of pages

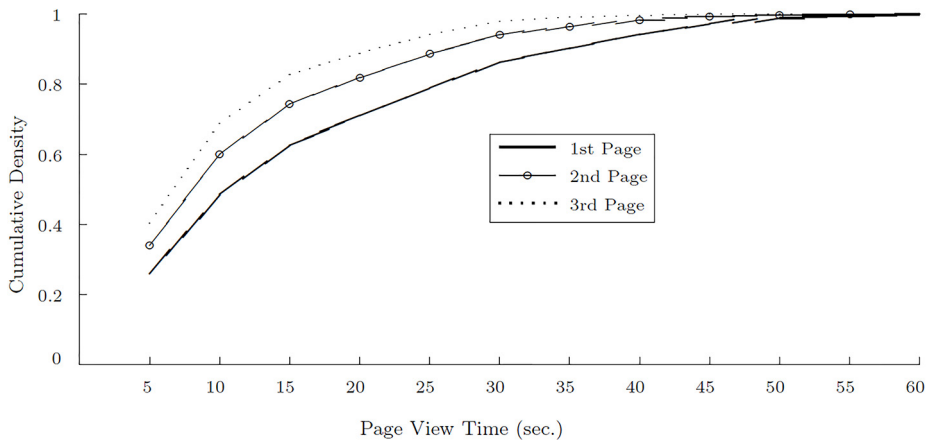


Table 3.
Model fit: revisit time

Model	AIC
μ	52,006
$\mu \sim \text{Gamma}(\theta, \delta)$	38,665
$\mu \sim \alpha \text{Gamma}(\theta_s, \delta_s) + (1 - \alpha) \text{Gamma}(\theta_i, \delta_i)$	35,881

supports our hypothesis on goal-oriented customers with shorter revisit durations and experiential customers with longer revisit durations.

5.3 Model validation

To assess the performance of the Markov chain models, we develop confusion matrices of the estimated models, as illustrated in Table 4. A confusion matrix allows displaying performance of a binary classification model. From a confusion matrix, we can derive performance measures of a model such as accuracy and precision.

Accuracy is the proportion of observations whose class the model can correctly predict.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

While accuracy is the most commonly used measure to evaluate the performance of a model, it may not be a good measure for imbalanced data sets. For example, when it comes to predicting customers’ booking probability, classifying all as negative (no booking) yields a 0.9775 accuracy score because the overall booking probability is 2.25%.

Precision is computed as the number of true positives over the number of positive predicted classes. It measures the exactness of a classifier. When the customers’ booking and revisit decisions are highly skewed to negative (no booking and no revisit), precision may be a better measure on how correctly the model predicts booking and revisit decisions.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Table 5 represents accuracies and precisions for the exit, booking and revisit probabilities. As a benchmark for precision, the overall percentage of the “Yes” class is also presented. For example, the percentage of customer visits that ends up with booking is 2.25% (in-sample) and 2.52% (out-of-sample). The precision values (7.66% for in-sample and 8.31% for out-of-sample) obtained from the model are more than three times higher than the overall booking probabilities. This means that, using the Markov chain model, we can improve the

	Predicted class	
	Yes	No
<i>Actual class</i>		
Yes	TP (True Positive)	FN (False Negative)
No	FP (False Positive)	TN (True Negative)

Table 4.
Confusion matrix

Decision probability	In-sample or out-of-sample	Accuracy (%)	Precision (%)	% of Yes class (benchmark to precision)
Booking, <i>b</i>	In-sample	82.74	7.66	2.25
	Out-of-sample	82.43	8.31	2.51
Revisit, <i>r</i>	In-sample	79.19	59.78	25.48
	Out-of-sample	79.73	62.70	26.34
Exit, <i>f</i>	In-sample	70.40	68.21	47.92
	Out-of-sample	69.20	66.90	47.62

Table 5.
Model evaluation

prediction power three times better. Overall, the suggested model increases prediction power for booking, revisit (62.7% precision vs 26.34% revisit probability) and search (66.9% precision vs 47.62% exit probability) decisions.

Furthermore, to validate the robustness of the model, we also compute out-of-sample performances. As shown in [Table 5](#), there are no significant differences between in-samples and out-of-samples, which confirm that the suggested model is robust.

6. Discussions

Empirical results provide us with important insights into online conversion behavior. First, higher discount rates increase not only purchasing probability but also customer engagement within the site (higher revisit and search probability). Price may have impacts on customers' loyalty throughout the whole decision-making process, from search to purchase. While many studies have explored how customers' search activities affect purchase conversions ([Moe and Fader, 2004](#); [Iwanaga et al., 2016](#); [Park and Park, 2016](#); [Mokryn et al., 2019](#)), very few studies have investigated the price effect on customer engagement within a website. Our empirical results imply that simple measure of search-to-purchase conversion rate may be misleading in the context of online promotion and marketing as price, often measured as discounts, may play a significant role in the purchase conversion. Thus, for proper prediction and evaluation of customer online activities, it is important to consider price factor. Moreover, price effects on purchase may not be linear as discount increases not only direct purchase probability but also enhances overall customer engagement throughout the entire purchase process. This non-trivial relationship between price and customer engagement may be potentially applied to developing a real-time marketing promotion tool based on customers' online activities.

Second, we found that search activities both within and across visits enhance overall consumer engagement, related to higher purchase and revisit propensities. Previous studies also suggested positive intra-visit search effects ([Zauberman, 2003](#); [Iwanaga et al., 2016](#)) and inter-visit search effects ([Bucklin and Sismeiro, 2003](#); [Johnson et al., 2004](#)) on customer engagement in online browsing. These findings lead to some practical implications for future studies and applications.

We have observed that revisiting customers have higher purchasing probability and higher with-in site engagement. If we are able to predict potential revisit customers, who have higher revisit probability, we may capture more latent demand and thus more revenue opportunities through promotional and/or customized marketing targeting the potential revisit customers. The empirical results presented in [Table 5](#) imply that it is possible to predict potential revisit customers with reasonable accuracy and precision. Prediction of revisit probabilities may be used for targeting potential buyers who are more likely to come back.

On the other hand, instantaneous probabilities for exit and search decisions can be used to detect customers who may end up with leaving without purchase. One potential application area is real-time marketing interventions through which firms may promote more search activities especially for those customers with relatively low level of search activities. Finally, our results indicate that online shoppers adjust their browsing durations both within and across visits. First, page view durations within a visit are shortened with additional page requests. This may imply self-imposed time constraints on online browsing or learning effects on page view times, which are consistent with previous studies including [Johnson et al. \(2004\)](#) and [Bucklin and Sismeiro \(2003\)](#). Second, our empirical results show that there may exist two distinct types of online browsing intentions ([Chatterjee et al., 2003](#); [Moe, 2003](#)): goal-oriented and experiential consumption. Consumers who come back after short durations are more likely to be goal-oriented and more determined to purchase. On the

other hand, customers who revisit after longer durations may be more experiential and may not commit immediate purchase. Based on this behavior, online booking websites may offer more differentiated marketing campaigns. For example, they may need to promote experiential customers (who have long inter-visit durations) to expand their search space by presenting other related products. For those who revisit after relatively short interval, they may provide tools and website design for easily narrow down searches and faster check-out to facilitate their goal-oriented (purchase-driven) behavior.

7. Conclusions

In this article, we have developed a model of airline search, revisit and booking activities of anonymous visitors. Our Markov chain approach provides a comprehensive framework to capture conversion probabilities as well as search and revisit durations. Using this stochastic model, we examined key research questions on online conversion behavior of anonymous visitors. Empirical studies based on airline booking clickstream data showed that low prices increase not only booking conversion rates but also overall site engagement. Moreover, positive search effects on site engagement are observed within a visit as well as across visits. Studies on inter-arrival times between searches provide an evidence for learning effects or time constraints on online shopping. Models for revisit durations support our hypothesis on two distinct revisit intentions – goal-oriented and experiential behavior. The research question of examining online conversion is complex, and several issues need to be further explored. For example, we were not able to include online activities across different sites. Also, the lack of customer information limited our research to clickstream analysis. Because of this limitation, our model does not include possible covariates such as customer demographics and customer perceptions on site design characteristics.

While our model is demonstrated on an airline booking website, it may be applicable to other areas in the tourism industry, in which understanding customers' online behavior is critical for capturing more revenue opportunities. Our stochastic model captures dynamic conversion of search, revisit and booking decisions as well as decision timing. However, online behavior may be different depending on site characteristics and type of products. Future extension of this model may include application-specific features of online search patterns.

References

- Bhatnagar, A., Sen, A. and Sinha, A.P. (2017), "Providing a window of opportunity for converting eStore visitors", *Information Systems Research*, Vol. 28 No. 1, pp. 22-32.
- Bucklin, R.E. and Sismeiro, C. (2003), "A model of web site browsing behavior estimated on clickstream data", *Journal of Marketing Research*, Vol. 40 No. 3, pp. 249-267.
- Cezar, A. and Ögüt, H. (2016), "Analyzing conversion rates in online hotel booking", *International Journal of Contemporary Hospitality Management*, Vol. 28 No. 2, pp. 286-304.
- Chatterjee, P., Hoffman, D.L. and Novak, T.P. (2003), "Modeling the clickstream: implications for web-based advertising efforts", *Marketing Science*, Vol. 22 No. 4, pp. 520-541.
- Chen, Y. and Yao, S. (2017), "Sequential search with refinement: model and application with click-stream data", *Management Science*, Vol. 63 No. 12, pp. 4345-4365.
- Iwanaga, J., Nishimura, N., Sukegawa, N. and Takano, Y. (2016), "Estimating product-choice probabilities from recency and frequency of page views", *Knowledge-Based Systems*, Vol. 99, pp. 157-167.
- Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S. and Lohse, G.L. (2004), "On the depth and dynamics of online search behavior", *Management Science*, Vol. 50 No. 3, pp. 299-308.

- Ko, H.-C. (2020), "Beyond browsing: motivations for experiential browsing and goal-directed shopping intentions on social commerce websites", *Journal of Internet Commerce*, Vol. 19 No. 2, pp. 212-240.
- Moe, W.W. (2003), "Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream", *Journal of Consumer Psychology*, Vol. 13 No. 1, pp. 29-39.
- Moe, W.W. and Fader, P.S. (2004), "Dynamic conversion behavior at e-commerce sites", *Management Science*, Vol. 50 No. 3, pp. 326-335.
- Mokryn, O., Bogina, V. and Kuflik, T. (2019), "Will this session end with a purchase? Inferring current purchase intent of anonymous visitors", *Electronic Commerce Research and Applications*, Vol. 34, p. 100836.
- Morosan, C. and Bowen, J. (2017), "Analytic perspectives on online purchasing in hotels: a review of literature and research directions", *International Journal of Contemporary Hospitality Management*, Vol. 30 No. 1, pp. 557-580.
- 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*, Vol. 13 No. 1, pp. 3-16.
- Park, C.H. and Park, Y.-H. (2016), "Investigating purchase conversion by uncovering online visit patterns", *Marketing Science*, Vol. 35 No. 6, pp. 894-914.
- Phocuswright (2017), "US market booking channels shifting", available at: www.phocuswright.com/Travel-Research/Research-Updates/2017/US-Market-Booking-Channels-Shifting (accessed 10 October 2019).
- Su, Q. and Chen, L. (2015), "A method for discovering clusters of e-commerce interest patterns using click-stream data", *Electronic Commerce Research and Applications*, Vol. 14 No. 1, pp. 1-13.
- Trusov, M., Ma, L. and Jamal, Z. (2016), "Crumbs of the cookie: user profiling in customer-base analysis and behavioral targeting", *Marketing Science*, Vol. 35 No. 3, pp. 405-426.
- Xie, K. and Lee, Y. (2019), "Hotels at fingertips: informational cues in consumer conversion from search, click-through, to book", *Journal of Hospitality and Tourism Technology*, Vol. 11 No. 1.
- Yeo, J., Hwang, S., Kim, S., Koh, E. and Lipka, N. (2020), "Conversion prediction from clickstream: modeling market prediction and customer predictability", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32 No. 2, pp. 246-259.
- Zauberman, G. (2003), "The intertemporal dynamics of consumer lock-in", *Journal of Consumer Research*, Vol. 30 No. 3, pp. 405-419.
- Zhang, Z., Liang, S., Li, H. and Zhang, Z. (2019), "Booking now or later: do online peer reviews matter?", *International Journal of Hospitality Management*, Vol. 77, pp. 47-158.

Further reading

- Lin, L., Hu, P.J.-H., Sheng, O.R.L. and Lee, J. (2010), "Is stickiness profitable for electronic retailers?", *Communications of the ACM*, Vol. 53 No. 3, pp. 132-136.

Corresponding author

Misuk Lee can be contacted at: leem@seattleu.edu