

An Information Stock Model of Customer Behavior in Multichannel Customer Support Services

Kinshuk Jerath

Columbia Business School, Columbia University, New York, New York 10027, jerath@columbia.edu

Anuj Kumar

Warrington College of Business Administration, University of Florida, Gainesville, Florida 32611, akumar1@ufl.edu

Serguei Netessine

INSEAD, Singapore 138676, serguei.netessine@insead.edu

We develop a model to understand and predict customers' observed multichannel behavior in a customer support setting. Using individual-level data from a U.S.-based health insurance firm, we model a customer's query frequency and choice of using the telephone or web channel for resolving queries as a stochastic function of her latent "information stock." The information stock is a function of the customer's "information needs" (which arise when customers file health insurance claims) and "information gains" (which customers obtain when they resolve their queries through the telephone and web support channels), and other factors such as seasonal effects (for instance, queries that arise at the time of annual contract renewal). We find that average information gain from a telephone call is twice as much as that from visiting the web portal; customers prefer the telephone channel for health event-related information but prefer the web portal for structured seasonal information; and customers are polarized in their propensities of using the web channel and can be broadly classified into "web avoiders" and "web seekers." Our model provides superior in-sample and out-of-sample fit than multiple benchmark models for aggregate and individual-level customer activity and has several managerial uses, such as capacity planning.

Keywords: multichannel customer behavior; customer service; call center; empirical operation management; probability modeling

History: Received: February 4, 2013; accepted: December 9, 2014. Published online in *Articles in Advance* April 27, 2015.

1. Introduction

Investment into after-sales customer support is crucial for customer satisfaction and, therefore, for customer retention and loyalty. The most visible example of this is a modern-day call center, which, on average, accounts for approximately 70% of business-to-consumer interactions at a firm (Mandelbaum 2006). Historically, customer service representatives at call centers responded to customers' queries by telephone. However, present day call centers offer a number of advanced technology-enabled support channels to respond to customers' queries. These support channels fall into two distinct categories: *assisted channels*, where the firm's representatives assist customers via telephone, email, short message service, and web chat; and *self-service channels*, where customers can find desired information via web-based self-service portals and interactive voice response (IVR) units.

Our main objective in this paper is to build a statistical model that can capture patterns in customers' observed behavior in a multichannel customer support setting. Subsequently, the estimated model can be

used as a decision support tool to predict future customer activity at the aggregate as well as individual levels, to evaluate the relative efficacy of the different customer support channels, and to gain various other insights into customers' usage of assisted and self-service support channels.

Firms deploy a variety of statistical models to predict and manage call center traffic (see Gans et al. 2003 for a review). However, most of the predictive models assume query arrival as an exogenous process and then model service time to allocate resources optimally. These models typically do not consider customer channel choice and the interactions among different channels. In this paper, we develop a novel information need-based framework to model channel choice in a multichannel customer support setting. The key idea behind this model is that customers use support channels to resolve queries that arise while using the product/service. In other words, a customer's *observed* channel usage behavior is driven by her *latent*, i.e., unobservable, information need. Note that we use the term "information need" to refer to anything that induces the customer to contact the

firm. For instance, in the context of a health insurance firm, a customer might want to understand the process of filing a claim for a certain type of medical service that she has used for the first time; this would classify as a query to gather information regarding an unfamiliar process. In another case, a customer might observe that the reimbursement she received after filing a claim was not as expected, and she might call the firm to resolve this issue; this would classify as a query for clarification. In both situations, the customer is calling the firm because she needs information about some details of her health insurance plan.

Our model is general enough to be applicable in a wide variety of scenarios. In this paper, we implement the model on data from the call center of a large U.S.-based health insurance firm that offers web- and telephone-based support to its customers. We model the latent “transactional information stock” for a customer at a given time as the composite of her information needs (which, in our context, arise when customers file health insurance claims) and information gains (which customers obtain when they contact the firm’s support channels to resolve their queries) up to that time. We assume that a customer’s observed channel usage behavior is a realization of a two-stage stochastic process—a query arrival process followed by a channel choice process—with the rates of the stochastic processes in both stages dependent on the current information stock of the customer. Besides the “transactional information stock” described above, we also account for a “seasonal information stock” to allow for queries that occur due to a seasonal event, such as renewing the insurance contract or changing the contract terms. We estimate the information stock model on individual-customer-level data on claims and channel usage.

We show that invoking the notion of an information stock is a good way to *model* observable consumer behavior because it is plausible and intuitive and, as we show empirically, the model’s performance is very good in terms of in-sample fit and out-of-sample predictions for total queries, telephone queries, and web queries. The model’s strong empirical performance provides support for our information stock-based modeling approach. Our model is also able to identify, with high precision, individual customers who are expected to have high probabilities of making telephone calls to the firm in the near future. This information can be valuable for the firm. For instance, the firm can proactively attempt to resolve the queries of such identified customers (say, by making calls to them in low traffic periods). This can help the firm to increase customer satisfaction as well as reduce some of its peak-time call volume, hence reducing customer service representative costs.

Furthermore, the parameter estimates from our model suggest that, in our setting, the telephone channel provides two times more information than the web channel. Thus, our model allows us to assess the customer-perceived values of different support channels from transactional data. We also account for heterogeneity across customers—we find a high degree of heterogeneity in query propensity, and we find a bipolar distribution of web choice probability, indicating the existence of two distinct customer segments: “web avoiders” and “web seekers.”

Our research makes several contributions to the literature. First and foremost, to the best of our knowledge, ours is the first attempt to take a multichannel customer support setting and model the customer query arrival and channel choice processes by modeling individual customers’ information needs and gains. This model provides highly accurate predictions of future query behavior at both the aggregate and individual levels. Second, our proposed framework utilizes information as a common denominator to understand the determinants of customers’ channel usage, with different sources contributing to this latent construct. This approach allows call center managers to provide a quantitative evaluation of the customers’ information demand and the firms’ information supply through different channels. Third, we provide a practical framework that allows a company, using limited transactional data on customer product and channel usage (usually captured in today’s business environment), to improve quality of service through better estimation of the query arrival process.

The rest of this paper is organized as follows. In §2, we discuss the related literature. In §3, we describe our research setting and our data, and a preliminary analysis to support our modeling approach. In §4, we develop our models, and in §5 we present our results. In §6, we conclude with the managerial implications of our research and outline future research directions.

2. Related Literature

Our work relates primarily to three streams of literature: customer behavior in a service support environment, policies for call center optimization, and multichannel customer behavior.

It is of great importance for firms to understand customer behavior in support services (Sousa and Voss 2006), which has motivated several papers on this topic. Bobbitt and Dabholkar (2001) and Meuter et al. (2005) explored the determinants of adoption and customer satisfaction for self-service technology (SST) channels using questionnaires and survey tools to elicit customers’ preferences regarding SSTs. However, they did not consider how adoption of SSTs affects demand for other available alternative channels. Xue et al. (2007) show that the adoption and

usage of various service channels (tellers, ATMs, and online banking) offered by a large retail bank depends on the demographic characteristics of customers. Campbell and Frei (2010) conducted a field study on the impact of online banking channel adoption on local branches, IVR, ATMs, and call centers. They show that users who adopted the online banking channel reduced their dependence on the IVR and the ATM (a *substitution* effect) but increased their consumption of the firm's call center and local branches (an *augmentation* or *aggravation* effect). Kumar and Telang (2012) conducted a controlled experiment to explain how customers' web portal usage affects their telephone usage. They find that in case of queries for which the web provides structured information, usage of the web leads to fewer telephone calls (the substitution effect), but in case of queries for which the web provides unstructured information, usage of the web leads to more telephone calls (the aggravation effect). In contrast, our primary purpose in the present study is to build a predictive model of customer multichannel behavior that can accurately predict the demand for different support channels offered by the firm at both the aggregate and individual levels. Aksin et al. (2013) focuses on customers' call abandonment decisions as influenced by their wait times, rather than on which customer service channel they use. In the present work, we address the question of customer channel usage by developing a probability model to estimate the relative efficacy of multiple support channels. Besides providing highly accurate predictions of the future activity of customers, our model also provides insights into customers' multichannel behavior.

Call center optimization is a highly researched topic in operations management. However, most of this work has been done primarily using analytical queuing models (Kleinrock 1975, Gans et al. 2003). There is a significant body of work on data-driven statistical models for predicting call center traffic to aid with staffing and workforce management decisions (e.g., Avramidis et al. 2004; Bassamboo and Zeevi 2009; Brown et al. 2005; Cezik and L'Ecuyer 2008; Mehrotra and Fama 2003; Mehrotra et al. 2010; Shen and Huang 2008; Soyer and Tarimcilar 2008; Taylor 2008, 2012; Weinberg et al. 2007). However, this literature typically models only telephone call arrivals and often assumes exogenous arrival rates for queries. In the present work, we model both query arrival and channel choice (with the telephone as only one of the various channels that consumers can use to make queries), and we model both of these processes as endogenous by making a customer's observed behavior a stochastic function of the customer's latent information stock, which is dependent on her history of health events and queries with the firm. Our

approach is related to the recent movement in operations management toward more accurately modeling consumer behavior in operational models (Netessine and Tang 2009). Additionally, there are related empirical studies that estimate the parameters of queuing systems in retail, healthcare, and fast food settings (Kim et al. 2015, Lu et al. 2013, Pierson et al. 2011), all of which, again, only model customers served by a single channel.

Recent advances in technology have enabled firms and customers to communicate via multiple channels; therefore, multichannel customer management has become one of the key challenges faced by practitioners (Neslin et al. 2006). Management of multichannel support services has become one of the hottest issues, as evidenced by industry surveys and trade publications (refer to http://www.egain.com/resources/white_papers/ and <http://multichannelmerchant.com>). Sun and Li (2011) study the interactions between onshore and offshore call centers. Marketing scholars have also studied issues in multichannel customer management, focusing primarily on the interactions among different sales channels such as online sales, physical store sales, catalog sales, etc. (see Ansari et al. 2008, Danaher et al. 2003, Deleersnyder et al. 2002, Geykens et al. 2002, Inman et al. 2004, Knox 2006, Shankar et al. 2003). In contrast, there is relatively little empirical work on multichannel customer support services. This research is an attempt to start filling this gap.

3. Research Setting, Data Description, and Preliminary Analysis

We study customer behavior at a multichannel call center of a major U.S. health insurance firm. The firm has a customer base of more than three million. Customers purchase annual health insurance plans from the firm and thereafter utilize the plans to get their medical expenditures reimbursed. During the health plan usage, customers often have queries regarding their plan coverage, status of claims, etc., for which they contact the firm. During the study period, the firm offered support to its customers via telephone (assisted channel) and the web portal (self-service channel). For web portal usage, customers have to first register at the web portal. Thereafter, they can visit the web portal at their convenience and obtain information regarding their plan benefits, their claim status, the details of participating health providers, general information on diseases, etc. Interacting with a customer via the web portal costs the firm an order of magnitude less than interacting via the telephone channel. This is in line with industry estimates that suggest that, per contact occasion, it costs a typical firm \$0.24 on average to interact with a customer via

the web portal, whereas it costs \$5.50 on average to interact via telephone (Kingstone 2006). Therefore, the firm in question is interested in accurately predicting future load on different channels for more efficient resource allocation and better understanding the determinants of customer channel choice.

We collected data for a random sample of 2,462 customers from the web-registered customer population of the firm. Roughly 35% of the firm’s three million customers were registered to use the web portal; these customers account for 64% of the total telephone calls to the firm’s call center. From our discussions with the firm’s representatives, we learned that customers who are heavier users of the insurance service are more likely to register on the web portal and also make significantly more queries—on average, a non-web-registered customer makes 0.53 telephone queries per year, whereas a web-registered customer makes 1.74 telephone queries and 3.13 web visits per year. Therefore, the firm is interested in learning the multichannel behavior of the web-registered customers. Although the results we present are valid only for the customers who use both the web and the telephone channels, our modeling framework is generic and can be applied to customers who have access to only the telephone channel.

For the 2,462 customers, we constructed an individual-level data set covering the 30-month time period from July 2005 to December 2007 by extracting relevant information from several disparate databases of the firm. Using the firm’s claims processing database, we collected data on the date of claim filing, the customer out-of-pocket expenses, and the provider charges for each claim. We extracted telephone usage information, specifically the date of a telephone call, from the call center’s automatic call distributor. Finally, we extracted web-portal-usage information from the firm’s web informatics database. Brief summary statistics on claims and queries are reported in Table 1.

The key idea behind our modeling approach is that filing claims leads to information needs for customers, and they resolve these needs by contacting support services with queries. This implies that, in our data, the queries should be correlated with the number and severity of claims. Before we develop the full model,

Table 1 Summary Statistics for 2,462 Customers for the Period from July 2005 to December 2007

Variable	Mean	Std. dev.	Median	Min	Max
No. of web queries	7.82	22.08	1	0	356
No. of telephone queries	4.36	5.38	3	0	42
No. of claims	78.26	69.5	60	0	725
No. of claims with customer out-of-pocket expenses	43.25	40.71	31	0	449

we test for this relationship by regressing the total monthly number of queries for customer i on the monthly number of claims she files and the provider charges in these claims:

$$Q_{it} = a_i + g_t + b_1 \text{CLM}_{it} + b_2 \text{CHRG}_{it} + \varepsilon_{it}, \quad (1)$$

where $i \in \{1, 2, \dots, 2462\}$ denotes the customer; $t \in \{0, 1, 2, \dots, 29\}$ denotes the months from July 2005 to December 2007; and Q_{it} , CLM_{it} , and CHRG_{it} denote, respectively, the total number of monthly queries, total number of monthly claims, and total monthly claim charges (in thousands of dollars). Customer-level fixed effects (a_i) account for unobserved differences across customers, and month-level fixed effects (g_t) control for seasonal variations in queries. For instance, customers are more likely to call in certain months of the year, such as during the insurance contract renewal month or during allergy seasons.

The estimate of b_1 is 0.020 and is statistically significant at the 1% level. This indicates a strong and positive correlation between the total number of monthly queries and the total number of monthly claims for customers. The estimate of b_2 is 0.004 and is statistically significant at the 5% level. This indicates that queries increase for claims with higher provider charges. After obtaining these preliminary reassuring results, we now proceed to develop our model.

4. Model Development

In this section, we first develop time series and probabilistic benchmark models, and then we develop our information stock-based model.

4.1. Model 1: Time-Series Benchmark Model

We first model customers’ queries with the following linear fixed-effect time series regression specification:

$$Q_{ijt} = \alpha_i + \gamma_t + \beta_1 \cdot \sum_{\tau=1}^{t-1} \text{CLM}_{i\tau} + \beta_2 \cdot \sum_{\tau=1}^{t-1} \text{CHRG}_{i\tau} + \beta_3 \cdot \sum_{\tau=1}^{t-1} Q_{ij\tau} + \varepsilon_{ijt}, \quad (2)$$

where $i \in \{1, 2, \dots, 2462\}$ denotes the customer, $j \in \{\text{telephone}, \text{web}\}$ denotes the channels of query, and $t \in \{0, 1, 2, \dots, 23\}$ denotes the months from July 2005 to June 2007. On the left-hand side of specification (2), Q_{ijt} denotes the total queries made by customer i on channel j in month t . On the right-hand side of specification (2), $\sum_{\tau=1}^{t-1} \text{CLM}_{i\tau}$, $\sum_{\tau=1}^{t-1} \text{CHRG}_{i\tau}$, and $\sum_{\tau=1}^{t-1} Q_{ij\tau}$ denote, respectively, the cumulative number of claims, claim charges, and queries of type j till month $t - 1$ for customer i . The parameters α_i and γ_t , respectively, capture the customer and month fixed effects. The customer fixed effects account for customer-level scale differences in queries, and the month fixed effects account for seasonality in queries.

4.2. Model 2: Probabilistic Benchmark Model

Following queuing theory-based work on call center management (Gans et al. 2003), we assume a Poisson query arrival process at the call center. Once the query arrives, the customer makes a Bernoulli choice between the telephone and web channel. Let λ_{i0} be the baseline mean query arrival rate and p_{i0} be the baseline web choice probability for customer i . We allow for seasonality in both the query arrival rate and the web choice probability. Accordingly, for customer i in month m , the mean query arrival rate is $\lambda_{im} = \lambda_{i0} \exp(I_m)$ and the web choice probability is $p_{im} = p_{i0} (\exp(\pi_s I_m) / [1 - p_{i0} + p_{i0} \exp(\pi_s I_m)])$, where I_m denotes 24 monthly parameters to be estimated $m \in \{0, 1, 2, \dots, 23\}$ (we use 24 months of data for estimation). The term π_s is a parameter that allows the impact of monthly indicator variables to depend on customers' channel choice probability and their query arrival rates. Let t_{ij} be the time of arrival of the j th query for customer i in month m , where $j = 1, 2, 3, \dots, x_i$ represents the sequence number of queries. The likelihood of the observed x_i query arrivals for customer i if her j th query comes in month m is

$$L_i = \prod_{j=1}^{x_i} \lambda_{im} e^{-\lambda_{im}(t_{ij}-t_{i(j-1)})} p_{im}^{y_{ij}} (1-p_{im})^{(1-y_{ij})},$$

where $y_{ij} = 1$ if the j th query for customer i is a web portal visit and 0 otherwise. For $i = 1, 2, \dots, N$ customers, the likelihood of the observed web and telephone queries for N customers is

$$L = \prod_{i=1}^N \left\{ \prod_{j=1}^{x_i} \lambda_{im} e^{-\lambda_{im}(t_{ij}-t_{i(j-1)})+I_m} p_{im}^{y_{ij}} (1-p_{im})^{(1-y_{ij})} \right\}.$$

In this model, λ_{i0} and p_{i0} are the latent propensities governing customer i 's observed behavior. Furthermore, different customers may have different latent propensities for their query processes. Some customers may inherently have the tendency to make more queries than others. Moreover, different customers may have different preferences about which channel to use—for instance, some web-savvy customers may prefer the web channel, whereas older customers may prefer the telephone channel. We allow for unobserved heterogeneity in customers' behaviors by allowing the baseline query arrival rates λ_{i0} to be distributed across customers as gamma distribution and baseline web choice probability p_{i0} as beta distribution. Specifically, we assume that $\lambda_{i0} \sim \text{gamma}(\gamma, \theta)$ and $p_{i0} \sim \text{beta}(a, b)$; i.e.,

$$f(\lambda_{i0} | \gamma, \theta) = \frac{\theta^\gamma \lambda_{i0}^{(\gamma-1)} e^{-\theta \lambda_{i0}}}{\Gamma(\theta)} \quad \text{and}$$

$$f(p_{i0} | a, b) = \frac{p_{i0}^{(a-1)} (1-p_{i0})^{(b-1)}}{B(a, b)}.$$

Note that we do not observe customer-level characteristics in our data and cannot control for observed heterogeneity. However, controlling for unobserved heterogeneity as explained above implies that we control for customer identities. We also note that the above model is equivalent to the negative binomial distribution (NBD)/beta binomial (BB) model with covariates (Fader and Hardie 2000, Goodhardt et al. 1984, Jeuland et al. 1980, Schmittlein et al. 1985).

4.3. Model 3: Information Stock Model

We wish to carefully model the query arrival and channel choice processes for customers. An effective approach to do this, used widely in the marketing and economics literatures, is to model a latent construct that drives observed behavior. For instance, McFadden (1973) introduced the idea that the latent construct, "utility," drives the stochastic choice process leading to observed consumer choice. Moe and Fader (2004) examine consumers' dynamic purchase behavior at an e-commerce website with a latent visit effect that evolves over visits. Netzer et al. (2008) examine alumni gift-giving behavior by using a non-homogeneous hidden Markov model in which donors transition from one latent relationship state with their university to another.

In a similar vein, we model the observed multi-channel behavior of a customer as a stochastic function of a latent information stock of the customer. We assume that, at a point in time, each customer has an "information stock" that determines her query frequency and channel choice behaviors. The information stock of a customer, in turn, is determined by the information needs that arise about the insurance contract as she uses the insurance plan (for instance, queries about medical coverage, claim processing procedures, and operational details of the contract) and the information gain that she obtains on contacting the firm through the telephone or the web channels. The model we build falls in the class of probability models of customer behavior (see Fader and Hardie 2009 for a review).

We categorize a customer's information stock into two broad categories. The first category is *transactional* information stock, which is determined by the health events faced by a customer. We assume that each insurance claim filed by the customer (or by a doctor's office on the customer's behalf) after a health event leads to information need C to the customer. In other words, with each additional claim, the information need of the customer increases by C . To fulfill her information needs, the customer can approach the firm through its support channels and receive some information gain from contacting the support channels. We assume that each web portal visit provides the customer with an information gain W , and each

telephone call provides the customer with an information gain T . These information needs and gains determine her transactional information stock. It is plausible that the information stock declines with time, i.e., a customer is more likely to make a query pertaining to more recent claims. Therefore, we allow for the decay of transactional information stock with time. Note that transactional information stock varies across customers based on the number and timings of claims they file and the queries they make; however, the values of C , W , and T are assumed to be the same for all claims, web visits, and telephone calls, respectively.

The claims filed and queries made by customers over a period can be organized into a set of sequences, where each sequence contains several claims followed by a query. Consider a customer denoted by i who receives n_{i0} claims $(1, 2, \dots, n_{i0})$ at time $t_{i01}, t_{i02}, \dots, t_{i0n_{i0}}$ and then makes the first query at time t_{i1} . We denote the net transactional information stock after the first query for customer i by I_{i1} , given by $I_{i1} = \sum_{v=1}^{n_{i0}} C\delta^{F(t_{i0v}, t_{i1})} - (y_{i1}W + (1 - y_{i1})T)$, where $y_{i1} = 1$ if the first query for customer i is a web portal visit and 0 otherwise, and $F(x_1, x_2)$ is a function that returns a natural number that denotes the number of months that have passed between calendar time x_1 and calendar time $x_2 > x_1$. We assume that transactional information stock decays at monthly intervals, i.e., it remains the same within a month but decays by a factor of δ for every calendar month that passes. This decay process for information stock is modeled on the lines of the decay process often used for “ad stock,” i.e., the cumulative lagged effect of advertising (Broadbent 1979, Danaher et al. 2008, Dube et al. 2005, Gijzenberg et al. 2011). Assume that this customer makes the j th query at time t_{ij} and then makes k claims after the j th query at time $t_{ij1}, t_{ij2}, \dots, t_{ijk}$. Between the two queries, several claims are filed by the customer. Let n_{ij} be the number of claims filed between the j th and $(j + 1)$ th query by customer i . We denote the net transactional information stock after k claims after the j th query for customer i by I_{ijk} , given by

$$I_{ijk} = \sum_{z=1}^j \left[\sum_{v=1}^{n_{i(z-1)}} C\delta^{F(t_{i(z-1)v}, t_{ijk})} - \{y_{iz}W + (1 - y_{iz})T\} \delta^{F(t_{iz}, t_{ijk})} \right] + \sum_{v=1}^k C\delta^{F(t_{ijv}, t_{ijk})}, \quad (3)$$

where $y_{iz} = 1$ if the z th query for customer i is a web portal visit and 0 otherwise. (For clarity, note that I_{ij} is the net stock of information need for customer i after her j th query and I_{ijk} denotes the net stock of information need for customer i after k claims after her j th query, i.e., between her j th and $(j + 1)$ th query.) Our

nomenclature implies that when a customer experiences an information gain, her transactional information stock reduces, and when she experiences an information need, her transactional information stock increases.

The dynamics in the transactional information stock are shown in Figure 1, using an example sequence of events for a customer. For this example, at the beginning of Month 1, the transactional information stock has the value 0. A claim arrives in the middle of the month at time t_{01} and the transactional information stock value becomes C . The value remains the same until time t_{02} , when another claim arrives and the transactional information stock value becomes $2C$. This value remains until time t_1 , when the customer makes a telephone query and the value becomes $2C - T$. When the second month starts, the transactional information stock decays by the multiplier δ , taking the value $\delta * (2C - T)$. The value remains at this level until time t_{11} , when another claim arrives and the transactional information stock value becomes $\delta * (2C - T) + C$. The process continues in this fashion, as shown in the Figure 1.

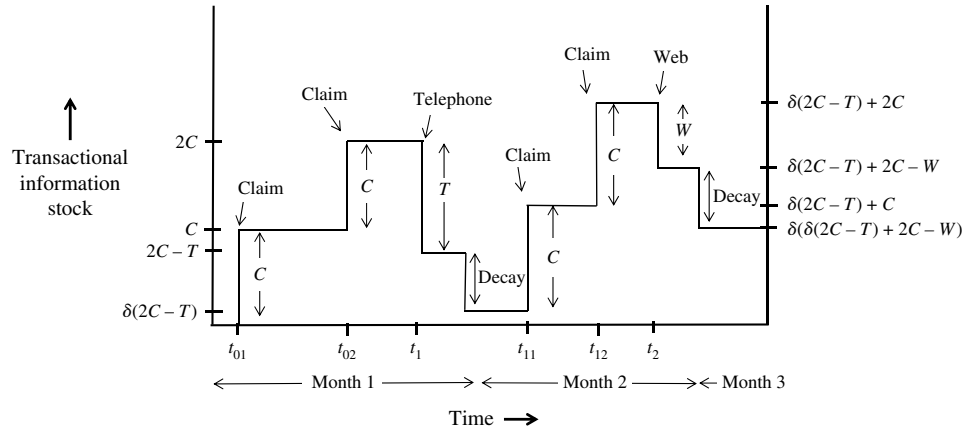
The second category of information stock is *seasonal* information stock, which is determined by the information needs and gains that arise from events relating to insurance plans at specific time periods. For instance, around the time of insurance contract renewal, customers make more queries regarding their ID card, renewal of their web portal login and password, insurance forms to be used in the upcoming year, etc. Similarly, the firm sends seasonal information bulletins to consumers, in allergy seasons, for instance, leading to information gains. These seasonal information needs are variable with time but equally applicable to all customers at a given point in time. We model seasonal information needs at the monthly level. We use parameters I_m to denote the seasonal information stock, in month m , where $m \in \{0, 1, 2, \dots, 23\}$. The total stock of information need for customer i after the k th claim after the j th query, and when the month is m , is given by $I_{ijk} + I_m$, and is the sum of the transactional and seasonal information stocks. We assume that customer query arrival and channel choice are driven by the total information stock that the customer has at a given time.

The rate for this process, for customer i after the k th claim after the j th query, and when the month is m , is given by

$$\lambda_{ijkm} = \lambda_{i0} \exp(I_{ijk} + I_m), \quad (4)$$

where λ_{i0} is the baseline mean query arrival rate for customer i . Therefore, the mean query arrival rate, λ_{ijkm} , will change for a customer with the arrival of claims or queries, or with a change in month; i.e.,

Figure 1 Illustration of the Dynamics in a Customer’s Transactional Information Stock



query arrival for a customer is modeled as a non-homogeneous Poisson process with the mean rate of arrival changing with the customer’s information stock. Note that a higher information stock denotes a higher information need, which corresponds to a higher query arrival rate.

Once the query arrives, the customer makes a Bernoulli choice with a web choice probability p between using the web and making a telephone call to resolve her query. Since telephone calls are answered by trained representatives of the firm, it is likely that when the information need is high, customers prefer to make a telephone call. Likewise, customers may prefer the web portal for structured information needs such as seeking insurance-contract-related information or applying for a new insurance card. We allow for these possibilities by modeling the web choice probability, p_{ijm} , as a function of the two types of information stocks. We define the web choice probability for the j th query for customer i , where the j th query for the customer arrives in month m , as

$$p_{ijm} = p_{i0} \frac{\exp(\pi_T I_{ij} + \pi_S I_m)}{[1 - p_{i0} + p_{i0} \exp(\pi_T I_{ij} + \pi_S I_m)]} \quad (5)$$

where p_{i0} indicates customer i ’s baseline web choice probability independent of information need. The term π_T is a parameter that allows the impact of transactional information stock to be different on customers’ channel choice probability and their query arrival rate. Similarly, π_S is a parameter that allows the impact of monthly indicator variables to be different on customers’ channel choice probability and their query arrival rate. The values of these parameters inform us of the sensitivity of the channel choice probability to the two kinds of information stock. Note that the Bernoulli web choice probability p_{ijm} for a customer changes with the arrival of claims and queries as well as with the change of month. (Note that we need channel choice probability only at the time of a query, and not after every claim, which

is why the index k , which denotes claims arriving between queries in λ_{ijkm} , does not appear in the channel choice probability in (5).)

We now develop the likelihood function for the observed data for customer i . For notational simplicity, we suppress the subscript m for month in the following derivation; i.e., we ignore the impact of the seasonal information stock and focus on the impact of the transactional information stock. Incorporating the seasonal information stock via month is straightforward since we only have to modify the query arrival rate and channel choice probability based on the month at a specific time.

The customer receives n_{i0} claims $(1, 2, \dots, n_{i0})$ at time $t_{i01}, t_{i02}, \dots, t_{i0n_{i0}}$ and then the first query arrives at time t_{i1} . The processes up to the first query of the customer are no query up to t_{i01} at rate λ_{i00} ; no query between t_{i01} and t_{i02} at rate λ_{i01} ; ...; no query between $t_{i0(n_{i0}-1)}$ and $t_{i0n_{i0}}$ at rate $\lambda_{i0(n_{i0}-1)}$; query arrival at t_{i1} with rate $\lambda_{i0n_{i0}}$ between $t_{i0n_{i0}}$ and t_{i1} ; and channel choice by customer for the first query with web choice probability p_{i1} . Therefore, the likelihood function for customer i up to the first query is

$$L_{i1} = e^{\sum_{u=0}^{n_{i0}-1} -\lambda_{i0u}(t_{i0(u+1)}-t_{i0u})} \times \lambda_{i0n_{i0}} e^{-\lambda_{i0n_{i0}}(t_{i1}-t_{i0n_{i0}})} \times p_{i1}^{y_{i1}} (1 - p_{i1})^{(1-y_{i1})},$$

where $y_{i1} = 1$ if the first query for customer i is a web portal visit and 0 otherwise.

The customer receives a total of x_i queries with n_{iz} claims between the z th and $(z+1)$ th query, where $z = 0, 1, 2, \dots, (x_i - 1)$. Customer i receives g_i claims after the x_i th query until the end of our period of observation (t_{end}), which is the same for all customers. The total likelihood function for customer i is

$$L_i = \prod_{z=0}^{x_i-1} [e^{-(\sum_{u=0}^{n_{iz}-1} \lambda_{izu}(t_{iz(u+1)}-t_{izu}))} \times \lambda_{izn_{iz}} e^{-\lambda_{izn_{iz}}(t_{i(z+1)}-t_{izn_{iz}})} \times p_{i(z+1)}^{y_{i(z+1)}} (1 - p_{i(z+1)})^{(1-y_{i(z+1)})}] \times e^{\sum_{u=0}^{g_i-1} -\lambda_{ix_i u}(t_{ix_i(u+1)}-t_{ix_i u})}, \quad (6)$$

where $t_{ix_i(g_i+1)} = t_{\text{end}}$. Total likelihood for $i = 1, 2, \dots, N$ customers is

$$L = \prod_{i=1}^N L_i. \quad (7)$$

As the subscript m for month was suppressed in (6), the Poisson mean query arrival rate λ_{ijk} and the Bernoulli web choice probability p_{ij} utilized in (6) and (7) are actually λ_{ijkm} and p_{ijm} , computed with appropriate transactional and seasonal information stock components as per Equations (4) and (5), respectively.

So far we have assumed that all claims for a customer give her the same information need C . However, claims with different characteristics may lead to different information needs. For instance, customers are more likely to make queries for claims where they have to pay out-of-pocket fees or for claims of higher value. Therefore, we allow for different information needs for claims based on whether the customer has to pay out of pocket. For customer i 's claim associated with health event h , we assume that

$$C_{ih} = C_0 \exp(\alpha_{LIAB} \cdot D_{LIAB, ih}), \quad (8)$$

where $D_{LIAB, ih}$ is a dummy variable that is equal to 1 if the claim has positive customer out-of-pocket expenses and 0 otherwise, α_{LIAB} is the impact of a claim with positive customer liability on the information need created by the claim, and C_0 is the baseline information need from a claim that is constant across all claims and across all customers.

Across customers, we assume gamma-distributed heterogeneity in the baseline query arrival rate, $\lambda_{i0} \sim \text{gamma}(\gamma, \theta)$, and beta-distributed heterogeneity in the baseline web choice probability, $p_{i0} \sim \text{beta}(a, b)$; i.e.,

$$f(\lambda_{i0} | \gamma, \theta) = \frac{\theta^\gamma \lambda_{i0}^{(\gamma-1)} e^{-\lambda_{i0}\theta}}{\Gamma(\gamma)} \quad \text{and}$$

$$f(p_{i0} | a, b) = \frac{p_{i0}^a (1 - p_{i0})^b}{B(a, b)}.$$

We account for the time of claim and query arrival in days in our information stock model (and the probabilistic benchmark model). This way we appropriately account for the actual time elapsed between events in our probability models, which is actually a duration model. However, we have operationalized two components of our information stock model at the monthly level: (1) the change in information stock due to seasonality and (2) the decay in transactional information stock. This is done primarily to reduce the complexity and computational burden of estimating the likelihood function.

We do not have data on the customers' demographic information, the content of their telephone

calls, or the web pages they visited. Our data are primarily on the "transactions" of the customer with the firm; these data are easy to collect and have no related privacy issues. Since most firms have ready access to such data, we expect our model to be useful for a typical firm's customer support center. Moreover, even with the limited data that we use, our model demonstrates excellent predictive power, as we show shortly.

Without the transactional information stock component, our model reduces to the probabilistic benchmark model, which is essentially the widely used NBD/BB model. Looked at another way, our model is the NBD/BB model augmented with the information stock component, the conceptualization and development of the latter being our key contribution. We also note that our model is quite general and can be applied to other similar situations as well. For instance, if there are more than two support channels, our model can easily be extended to a NBD/Dirichlet model (Goodhardt et al. 1984) with straightforward adjustments to the expressions for information stocks. Similarly, our model can be easily estimated for non-web-registered customers who have access to only one channel, i.e., the telephone channel. Necessary tweaks can also be made to the query-arrival process if needed to capture patterns in the data at hand; for instance, instead of the Poisson arrival process, the more flexible Erlang-2 arrival process can be used for query arrival (Jeuland et al. 1980). In our specific case, the Poisson arrival process is sufficient.

5. Estimation and Results

5.1. Estimation Procedure

We estimated the parameters of Model 1 using the fixed-effect ordinary least squares regression routine in STATA. To estimate the parameters of Models 2 and 3, we used a hierarchical Bayes framework (Gelman et al. 2009). For the information stock model (Model 3), we grouped parameters into two sets: (1) information stock-related parameters ($C_0, W, T, \delta, \alpha_{LIAB}, I_m, \pi_T, \pi_S$) and (2) parameters that determine heterogeneity in baseline rates across customers (γ, θ, a, b). For our probabilistic benchmark model (Model 2), we only had the heterogeneity parameters ($\gamma_T, \theta_T, \gamma_W, \theta_W$). We used a Metropolis-Hastings algorithm to successively draw from the following Markov chain Monte Carlo (MCMC) chains: (1) draw($\lambda_{i0}, p_{i0} | C_0, W, T, \delta, I_m, \pi_T, \pi_S, \alpha_{LIAB}, \gamma, \theta, a, b, \text{data}_i$); (2) draw($C_0, W, T, \delta, I_m, \pi_T, \pi_S, \alpha_{LIAB} | \gamma, \theta, a, b, \lambda_{i0}, p_{i0}, \text{data}_i$); and (3) draw($\gamma, \theta, a, b | C_0, W, T, \delta, I_m, \pi_T, \pi_S, \alpha_{LIAB}, \lambda_{i0}, p_{i0}$). Detailed algorithms for the MCMC chain are provided in the online appendix (available as supplemental material at <http://dx.doi.org/10.1287/msom.2015.0523>).

We ran 40,000 iterations of the MCMC steps; the first 30,000 iterations were used as initial burn-in to reach convergence, which we checked visually; the last 10,000 iterations were used to infer the posterior distributions of the parameters. We used multiple starting values for the MCMC chains and confirmed that the parameters converged to the same values.

5.1.1. Identification and Parameter Recovery. We now briefly discuss how the parameters in the proposed model are identified given the variation in our data. The parameters γ and θ are identified by the differences in overall mean query arrival rates across customers. Similarly, the parameters a and b are identified by the differences in overall web choice rates across customers. In addition, the parametric forms that we assume for these heterogeneity distributions also help in identification. The seasonal information stock parameters I_m are identified by the common variation in query rates and web choice rates across calendar months for all customers. The transactional information stock parameters C_0 , W , and T are identified by the variation in query rates at different channels for different periods with respect to the average overall claim arrival and query rates for customers in that period, after controlling for the baseline query rates and common monthly variations in query rates across customers. The decay parameter δ is identified by the differences in customers' actions in periods immediately after a sequence of claims versus actions several periods after. The parameter α_{LLAB} is identified by the variation in the presence of out-of-pocket expenses across different claims. The parameters π_T and π_S are identified by the variation in web choice probabilities with the changes in the two categories of information stocks, after controlling for overall web choice probabilities.

To further check that parameters of our model are well identified, we conducted a simulation study. In this study, we simulated data using sets of predetermined values of parameters in our model such that we covered a variety of cases for different relative information values of the different support channels and the heterogeneity in latent propensities across customers. Then, using the procedure described above, we estimated the model on the simulated data to check (i) whether the recovered parameter values match the actual parameter values used for data generation and (ii) whether the estimated parameter values can accurately recover aggregate query volumes in the data. We find that, in all the cases we considered, the recovery of parameters as well as of aggregate statistics in the data is very good. This analysis provides confidence in our estimated parameters. More details are available in the online appendix.

Table 2(a) Estimation Results for Model 1 (Times-Series Benchmark Model)

Coefficient estimate (cluster corrected robust standard error)	Telephone queries	Web queries
β_1 (impact of cumulative lagged claims)	0.002*** (0.000)	0.002*** (0.000)
β_2 (impact of cumulative lagged provider charges)	-0.003 (0.003)	-0.006 (0.005)
β_3 (impact of cumulative lagged queries)	-0.046*** (0.002)	-0.001** (0.000)
R^2	0.71	0.64

*** and ** denote statistically significant coefficient estimates at the 1% and 5% levels, respectively.

Table 2(b) Estimation Results for Models 2 and 3

	Probabilistic benchmark model (Model 2)	Information stock model (Model 3)
In-sample fit statistics		
Log marginal density	-112,138.41	-110,862.89
Log Bayes factor		1,275.52
Parameter estimates		
γ	0.683 [0.638, 0.719]	0.887 [0.835, 0.926]
θ	64.691 [58.526, 70.699]	126.16 [112.02, 136.13]
a	0.490 [0.456, 0.526]	0.492 [0.448, 0.532]
b	0.701 [0.651, 0.777]	0.716 [0.653, 0.798]
C_0 (baseline information need from claim)		0.258 [0.209, 0.305]
W (information gain from web visit)		0.835 [0.795, 0.871]
T (information gain from telephone call)		1.801 [1.679, 1.935]
δ (decay factor)		0.425 [0.406, 0.447]
π_T (impact of transactional information stock on query channel choice)		-0.039 [-0.158, -0.077]
π_S (impact of seasonal information stock on query channel choice)	0.302 [0.164, 0.432]	0.303 [0.154, 0.474]
α_{LLAB} (impact on information need for claim with customer liability)		-0.007 [-0.250, 0.214]

Note. For each parameter, we report the posterior mean followed by the 95% credible interval.

5.2. Model Estimates

We calibrated Models 1–3 for the first 24 months of data (from July 2005 to June 2007) and used the last six months of data (from July 2007 to December 2007) as a holdout sample.

In Table 2(a), we report the estimates for the time series benchmark model (Model 1). These estimates show that queries (both telephone and web portal)

are positively correlated with the lagged cumulative number of claims and negatively correlated with the lagged cumulative number of queries.

In Table 2(b), we report the estimates from the probabilistic benchmark model (Model 2) and the information stock model (Model 3). We report the values of the 23 seasonal information stock parameters for Model 3 in Table A2 in the online appendix. In Table 2(b), the higher value of the log marginal density for the information stock model and the large value of the log Bayes factor suggest that the information stock model fits the observed data better than the probabilistic benchmark model.

5.3. Model Predictions

Since the main objective of our model is to accurately predict the future queries at different channels, we rigorously analyze its aggregate-level and individual-level predictive accuracy.

5.3.1. Aggregate-Level Predictions. For Models 1–3, we predict total queries, telephone queries, and web queries for each customer in our sample for the calibration period (from July 2005 to June 2007) as well as the hold-out period (from July 2007 to December 2007). To test these predictions from the models, we aggregate them across the full cohort of customers and across time to the monthly level. We report the mean absolute percentage error (MAPE) values for the in-sample and out-of-sample predictions from the three models in Table 3. It is clear from this table that, as compared to the benchmark models, the information stock model makes significantly superior in-sample and out-of-sample predictions for total queries, telephone queries, as well as web queries.

Our model could be used by the firm for capacity planning for the telephone and web channels. An aspect that is important in capacity planning is overpredictions and underpredictions by the model. Whereas overprediction of calls results in excess deployment of customer service representatives (CSRs) and hence extra costs, underprediction leads to shortage of CSR deployment and consequently higher call blockage/abandonment rates and call waiting, which lead to higher levels of customer dissatisfaction. To assess this for each model, we make monthly out-of-sample predictions (i.e., for the

months from July 2007 to December 2007) for telephone queries and web visits and track whether the model overpredicted or underpredicted for a particular month. We then assign different importance weights to the overpredictions and underpredictions and find a consolidated error percentage for the different models, averaged across the six months. In Table 4, we show the error numbers for different sets of weights for over- and underprediction. Note that equal weightage of overpredictions and underpredictions in the first row yields the same error numbers as shown in Table 3; the second row assumes that overpredictions are twice as costly as underpredictions, and the third row assumes that underpredictions are twice as costly as overpredictions. In each case, we find that the information stock model does significantly better than the benchmark models for both telephone and web predictions. We have considered other relative weights of over- and underprediction errors as well, and this pattern holds consistently. Overall, we can say that for the purposes of capacity planning, our proposed model will perform significantly better than the benchmark models.

5.3.2. Individual-Level Predictions. An additional advantage of the information stock model is that it can predict queries at the individual customer level based on each customer’s calculated information stock, using the customer’s claim and contact history. Thus, the information stock model should make more accurate predictions of queries at the individual customer level as compared to the benchmark models. We can use the information stock model to better identify, as compared to the benchmark models, the customers likely to make a telephone call in a particular time period (the “calling customers”). We compare the predictive power of the information stock model and the benchmark models in identifying individual calling customers in two ways.

First, we use the information stock model to compute the calling probability for each customer in the out-of-sample period (from July 2007 to December 2007). We then sort customers in descending order of their calling probabilities given by the model. We do the same for the benchmark models. We designate the customers with high calling probability as the “calling customers.” To assess the predictive power of our model at the individual customer level, we

Table 3 Prediction Errors (MAPE)

	In-sample predictions (%)			Out-of-sample predictions (%)		
	Total queries	Telephone queries	Web queries	Total queries	Telephone queries	Web queries
Times-series benchmark model (Model 1)	7.30	7.70	9.15	18.77	18.10	26.41
Probabilistic benchmark model (Model 2)	5.91	8.24	9.50	11.23	10.71	20.10
Information stock model (Model 3)	2.40	5.11	3.22	4.75	5.10	10.32

Table 4 Out-of-Sample MAPE with Different Weights for Over- and Underprediction

Over:Under weight	Time-series benchmark model (%)		Probabilistic benchmark model (%)		Information stock model (%)	
	Telephone queries	Web queries	Telephone queries	Web queries	Telephone queries	Web queries
1:1	18.10	26.41	10.71	20.10	5.10	10.32
2:1	16.58	30.06	8.57	22.49	5.79	10.58
1:2	19.07	22.76	12.43	18.19	4.24	10.15

Table 5 Identification of Calling Customers

Top percentile of customers based on their calling probabilities (%)	% correctly identified calling customers		
	Time-series benchmark model	Probabilistic benchmark model	Information stock model
25	40.13	57.93	85.39
50	60.72	75.19	94.33
75	84.82	90.79	98.29
100	100.00	100.00	100.00

compare how many of the calling customers identified by each model actually called in the out-of-sample period (i.e., from July 2007 to December 2007). In Table 5, we show this metric for different cumulative quartiles of customers (the quartiles are based on their computed calling probabilities). Therefore, the first row of the table implies that for the top quartile of customers identified as calling customers by the time-series benchmark model, 40.13% actually called in the out-of-sample period; for the top quartile of customers identified as calling customers by the probabilistic benchmark model, 57.93% actually called in the out-of-sample period; and, for the top quartile of customers identified as calling customers by the information stock model, 85.39% actually called in the out-of-sample period. This shows that the information stock model identifies calling customers quite well, and vastly more accurately than the benchmark models. (If we consider the top 100% of customers sorted based on the calling probabilities (i.e., if we consider *all* the customers), all calling customers are identified.)

The analysis above helps to determine the accuracy of the different models in identifying the top calling customers. Second, we conduct alternative rigorous analysis of the model's predictive ability. Specifically, we sort calling customers by the number of *actual* queries and then segment the cohort into quartiles by the degree of activity, and then see the performance of the models across these quartiles. This ensures that we compare the predictive accuracy of the three models on the same cohort of calling customers. In Table 6, we provide the correctly predicted calling customers in each quartile from each model in each month of the out-of-sample period. Note that quartile 1 is composed of the highest-calling customers and quartile 4 of the lowest-calling customers. The averaged numbers in the last row of the table clearly show that the information stock model performs the best in terms of identifying calling customers.

So far we have compared the model's accuracy in correctly identifying calling customers. However, to assess the overall predictive power of our model, we need to compare its accuracy in identifying both calling and noncalling customers with the benchmark models. Therefore, we computed rates of true positives (TP; correctly predicted calling customers), false positives (FP; incorrectly predicted calling customers), true negatives (TN; correctly predicted noncalling customers), and false negatives (FN; incorrectly predicted noncalling customers) from the different models. For each month, we predict the calling and noncalling customers from each model to determine the rates of TP, FP, TN, and FN.

In Table 7, we provide the percentages of TP, TN, FP, and FN, averaged over the six months. From Table 7, we find that the information stock model does significantly better in correctly identifying the actual calling customers (86.6%) as compared to the benchmark models (66.2% and 47.7% for the probabilistic and time-series benchmark models, respectively). Moreover, the information stock model also makes significantly fewer errors in misidentifying the noncalling customers as calling customers (10.7%) as compared to the benchmarks (27.3% and 40.2% for

Table 6 Quartile-Wise Accuracy in Identifying Calling Customers

Month	Actual calling customers per quartile	Number of correctly predicted calling customers											
		Time-series benchmark model				Probabilistic benchmark model				Information stock model			
		1st qrt	2nd qrt	3rd qrt	4th qrt	1st qrt	2nd qrt	3rd qrt	4th qrt	1st qrt	2nd qrt	3rd qrt	4th qrt
July	69	41	38	32	29	55	51	48	37	66	63	59	57
August	72	38	34	31	25	58	48	36	30	70	66	60	57
September	62	35	31	25	22	48	41	37	36	58	55	50	46
October	69	38	32	27	24	52	46	43	44	64	62	59	54
November	63	37	32	25	24	47	43	41	39	58	57	54	52
December	52	36	30	24	22	40	35	32	32	47	46	42	37
Average % correctly identified		58.6	58.6	51.2	42.5	37.9	68.2	61.4	56.8	93.6	93.6	90.1	83.6

Table 7 Average True Positive, False Positive, True Negative, and False Negative Rates

	Time-series benchmark model (%)	Probabilistic benchmark model (%)	Information stock model (%)
True positive	47.7	66.2	86.6
False positive	40.2	27.3	10.7
True negative	95.3	96.8	98.7
False negative	6.1	4.0	1.5

the probabilistic and time-series benchmark models, respectively). This clearly shows the superior performance of the information stock model in identifying calling and noncalling individuals.

Identification of calling customers with high accuracy can be a great advantage to a firm. For instance, once the high-calling-probability customers are identified, to resolve their queries proactively and preempt some of their calls, the firm’s CSRs can reduce peak-time calls by making outgoing calls to them at nonpeak times when the CSRs are free. As CSRs are deployed at call centers primarily based on the predicted peak-time call load, making calls to customers in advance may reduce the peak-time call load and thus save CSR-related costs for the company. For instance, according to Table 5, if the top 25% of customers (based on calling probability) are considered, the information stock model, on average, is able to correctly identify approximately 85.39% of the calling customers in a month. If 30% of the total customers make calls during peak times, then approximately $85.39 \times 0.3 = 26\%$ of the total customers making peak-time calls would be correctly identified by our model. If we assume that by making outgoing calls to these customers, calls from half of these customers are avoided—i.e., approximately $26/2 = 13\%$ of total calls are avoided—significant cost savings for the firm would ensue. For instance, the firm from which we obtained data has approximately three million customers making on average 440,000 calls per month; in this case, making outgoing calls to the identified customers may result in a reduction of roughly

$440,000 \times 0.13 = 57,200$ peak-time calls per month, or $57,200/22 = 2,600$ peak-time calls per day (assuming 22 working days per month), which would lead to significant savings.

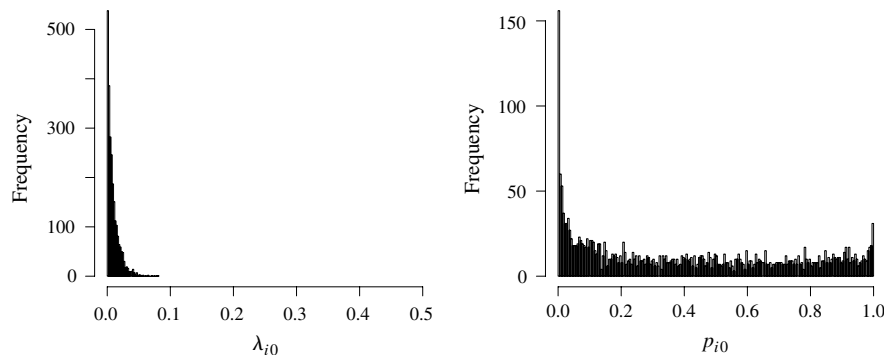
Note that call centers typically make staff allocation plans at the weekly, daily or, sometimes, even hourly levels. Even though our analysis is conducted at the monthly level, the results obtained can be used to make decisions for shorter time frames. For instance, once a planner is given a list of customers with a high probability of calling in the near future (specifically, in the next one month), he or she can use this information to make scheduling and staff allocation decisions for time frames shorter than one month as needed. Naturally, if we had more data, we could model information at the weekly or even daily levels, which may improve the performance of the model.

5.4. Inferences from the Information Stock Model

In this section, we discuss inferences from the estimated parameter of the information stock model. The distribution of the mean query arrival rate across customers in our sample is shown in the left plot of Figure 2. This long-tailed plot indicates large heterogeneity in the baseline query arrival rates—a majority of the customers have a low query rate while a few customers have the propensity to make a large number of queries. The median number of queries is 2.27 per year. The distribution of the web choice probability across customers in our sample is shown in the right plot of Figure 2 and indicates a polarized distribution; i.e., some customers (a relatively larger number) have quite low web choice probability and can be classified as “web avoiders,” whereas other customers (a relatively smaller number) have quite high web choice probability and can be classified as “web seekers.” The median web choice probability is 0.36.

The estimated value of the baseline information need from a claim, C_0 , is 0.258; of the information gain from a web visit, W , is 0.835; and of the information gain from a telephone call, T , is 1.801. This implies that a telephone call provides the information gain equivalent to the information need generated by

Figure 2 Distribution of Baseline Query Arrival Rate and Web Choice Probability



$1.801/0.258 = 6.98$ claims; a web query provides the information gain equivalent to the information need generated by $0.835/0.258 = 3.24$ claims, and a telephone call provides the information gain equivalent to $1.801/0.835 = 2.16$ web visits. Thus, in our setting, a telephone call, on average, provides a larger information gain than the web channel, and it is therefore significantly more effective in resolving queries. However, average industry estimates suggest that on average it costs firms approximately US\$0.24 to respond to a customer's query on the web and US\$5.50 to respond through the telephone channel via a call center (Kingstone 2006). Therefore, the information gain per dollar for the web channel is $0.851/0.24 = 3.55$, whereas the corresponding value for the telephone channel is $1.801/5.5 = 0.33$. In other words, the information gain per dollar is more than 10 times higher for the web channel than for the telephone channel.

Next, we examine the impact of the two types of information stocks on the probability that a customer uses the web channel to resolve a query. The negative and significant estimate for π_T (-0.039) suggests that the probability of web usage decreases with higher transactional information stock. Since transactional information stock is generated from health events, this implies that customers prefer to use the telephone channel for information needs generated by health events. In contrast, the positive and significant estimate for π_S (0.303) suggests that the probability of web usage increases with higher seasonal information stock. As pointed out earlier, seasonal information needs are typically related to the insurance contract and include requests for ID cards, password/login updating, etc. This category of information is easy to retrieve on the web portal, and thus customers tend to use the web portal for obtaining this type of information. These results are also in line with the results of Kumar and Telang (2012), who also studied a health insurance setting and found the web to be effective in resolving structured queries, but found the telephone to be effective in resolving unstructured and complex queries.

We find a significant estimate for δ (0.425), which translates into approximately 57.5% decay of transactional information stock in a month. We did not find the estimate of α_{LAB} to be statistically significant in our data.

It is important to note that the model allows for stochasticity in customer behavior—larger λ_{ijkm} implies that there is a *tendency* to make queries faster, and larger p_{ijm} implies that there is a larger *tendency* to choose the web channel. Therefore, as more claims arrive for a customer, information needs increase. The customer could make a query even with a small information need that is generated when a claim arrives but is more likely to make a query as the information

need increases. A larger information need also makes it more likely that the telephone channel will be used.

Taking an overall view, the insights obtained from our parameter estimates lend face validity to our modeling approach. Note that the values above are specific to this setting and depend on many factors, such as the level of training of the CSRs who take calls and the helpfulness and ease-of-use of a company's website. Needless to say, these values may differ in other settings. The model is flexible enough to capture different patterns in other data sets but would have to be reestimated to be informative for those settings. In the present study, we have conceptualized information as a one-dimensional construct. One can, however, think of information as a multidimensional construct; e.g., one dimension can capture complicated/unstructured information needs and another dimension can capture simpler/structured information needs. It may then be the case that the web channel is effective for simple/structured information needs but for not for complicated/unstructured information needs, whereas the telephone channel is effective for both. This is suggested by the result that if information needs emanate from the seasonal information stock, then consumers tend to prefer the web channel. However, to fully develop and estimate a model with multiple dimensions of information, we would need more details of the consumers' interactions with the firm for every query instance (e.g., transcripts of telephone conversations, and clickstream data for the web visits).

5.5. Robustness Checks

We conducted the following robustness checks and found no qualitative differences in the parameter values and insights obtained: (1) we calibrated the model with the first 18 months of data as well as the full 30 months (instead of 24 months of data) and obtained similar parameter estimates; (2) we estimated a model with 11 variables for I_m to capture seasonal information stock (one for each calendar month across years, i.e., the same parameter for July 2005 and July 2006, for August 2005 and August 2006, and so on), and obtained similar estimates to those reported here for the other parameters; (3) we checked for possible learning by consumers for the web channel by allowing for higher information gains in later web visits using a multiplicative factor and did not find the learning effect as significant; and (4) we estimated our model under the assumption that a telephone call results in complete resolution of the customer's questions—i.e., after a telephone call, the customer's information stock becomes zero and found inferior model fit on the data as compared to our information stock model.

In addition to the above, we used particular features of the data on healthcare provider charges to

code health events as repeat health events (e.g., events associated with multiple claims for a chronic disease). The key idea is that information needs are low in the case of repeat events because the consumer would already have most of the relevant information from previous experience. We find that though there is an improvement in the in-sample and predictive performance of every model, this improvement is small.

Finally, we estimated the information stock model by considering only the information need from claim arrivals and information gain from telephone calls; i.e., we ignore the data on web visits in this estimation. We find that this model has considerably worse performance than the full information stock model that uses the web data (details are available in the online appendix). This underscores the value of information on customers' web usage in predicting their multichannel behavior.

6. Conclusions and Future Work

In this paper, we propose a novel information stock-based framework to endogenously model the query generation and channel choice processes of customers in a multichannel customer support services setting. We implement the proposed model on individual-customer-level data obtained from a large US-based health insurance firm. We find that the information stock model can accurately capture patterns in customers' multichannel query behavior. The model provides accurate predictions for aggregate query volumes for the different support channels. Furthermore, it is able to identify with high accuracy the customers who are likely to make queries in the near future. Therefore, the model can serve as a useful managerial tool. For instance, making advance outgoing calls to customers who have high calling probabilities can help reduce peak-time calls, which can lead to substantial cost savings for the firm.

We also find that, in our setting, the average information gain from a telephone call is slightly more than double the information gain from visiting the web portal; i.e., the telephone channel, on average, is significantly more effective than the web channel in resolving customers' queries. Regarding channel choice, we find that customers prefer the telephone channel for higher health-event-related information needs and they prefer the web portal for more structured seasonal information needs. We also find that there is a large degree of heterogeneity in customers' propensity to use the web portal while making a query—some customers are "web avoiders" and others are "web seekers."

Our results may be applicable to the specific situation that we study, but our modelling framework is general enough for use in other settings with straightforward adjustments and can inform managers about

the relative efficacies of their customer support channels. For instance, the web portal can provide lower value than the telephone in complex services (such as health insurance), but may provide a higher value in less complex services (such as personal banking). Comparative studies of this kind, where the value of different types of customer service channels is determined for different industries and for different firms in an industry, would be an excellent application of this model and an avenue for future work.

Our results have significant managerial implications for new-generation multichannel customer service operations technologies. At a time when web portals are becoming a popular choice for reducing customer service costs, our estimates indicate the superior informational value of the traditional telephone channel over the self-service web portal. This suggests that the assisted telephone channel is still a dominant customer support channel at least for complex services such as health insurance. Our estimates also suggest that web portals are effective for simple, unambiguous tasks (such as seasonal information needs). We expect our results to extend to other self-service support channels. Therefore, managers have to make a careful choice regarding what infrastructure to set up for customer support services. The design of the web portal, in terms of ease of access of information, can be an important dimension in this decision.

Our model can be extended in many different directions in the future. First, we have conceptualized information as a one-dimensional construct. It may be useful to think of information as a multidimensional construct; e.g., one dimension can capture structured information needs and another can capture unstructured information needs. Second, details about the nature of health events would allow us to more precisely model information needs for different types of claims. Third, currently we treat the claim arrival process for customers as exogenous, but this process could be endogenous. For instance, based on how the insurance firm's representatives treat the customer, she may change her claim frequency (e.g., in an extreme response to unsatisfactory query resolution, she may switch the insurance company and her claim frequency will become zero). Future work could explore the endogeneity between claim frequency and query behavior. Fourth, we studied two channels in the model, namely telephone and web. Future work can extend the model to include more customer support channels (such as short message service and email channels) as needed; this would be a fairly straightforward extension.

Finally, we have timing data on the claims and queries of customers and some data on claim characteristics. On one hand, this implies that our model is

useful for the typical firm, since these data requirements are not heavy and the model still generates accurate predictions and useful insights. On the other hand, these data limitations present an opportunity to further enrich the model by extending it to incorporate more data. Richer data can allow us to add heterogeneity to other model components, such as the information gains from different web visits and telephone calls and the information needs from different claims. This would allow the possibility, for instance, that some web visits lead to an information need rather than a gain (say, because of confusion after obtaining online information), with the implication that a future query is accelerated rather than delayed. Additional data such as call transcripts and clickstream data could allow us to model such dependencies across queries. For instance, from the web pages visited by a customer, we could determine whether or not the customer was able to obtain the information she sought; if not, a subsequent query may actually be accelerated and the customer may be more likely to use the telephone channel.

Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/msom.2015.0523>.

Acknowledgments

The authors are listed in alphabetical order and have contributed equally. They thank the Marketing Science Institute and the Wharton Customer Analytics Initiative for generous support for this research. They also thank seminar participants at INSEAD, Indian School of Business, the Cross-platform and Multichannel Consumer Behavior conference organized by the Wharton Customer Analytics Initiative in 2010, and the Challenges and Opportunities in Managing IT Enabled Multichannel Operations conference at the University of Florida in 2012.

References

- Aksin Z, Emadi S, Ata B, Su C-L (2013) Structural estimation of callers' delay sensitivity in call centers. *Management Sci.* 59(12):2727–2746.
- Ansari A, Mela CF, Neslin SA (2008) Customer channel migration. *J. Marketing Res.* 45(1):60–76.
- Avramidis AN, Deslauriers A, L'Ecuyer P (2004) Modeling daily arrivals to a telephone call center. *Management Sci.* 50(7):896–908.
- Bassamboo A, Zeevi A (2009) On a data-driven method for staffing large call centers. *Oper. Res.* 57(3):714–726.
- Bobbitt LM, Dabholkar PA (2001) Integrating attitudinal theories to understand and predict use of technology-based self-service: The Internet as an illustration. *Internat. J. Service Indust. Management* 12(5):423–450.
- Broadbent S (1979) One way TV advertisements work. *J. Market Res. Soc.* 23(3):139–166.
- Brown LD, Gans N, Mandelbaum A, Sakov A, Shen H, Zeltyn S, Zhao L (2005) Statistical analysis of a telephone call center: A queuing science perspective. *J. Amer. Statist. Assoc.* 100(469):36–50.
- Campbell D, Frei FX (2010) The cost structure, customer profitability, and retention implications of self-service distribution channels: Evidence from customer behavior in an online banking channel. *Management Sci.* 56(1):4–24.
- Cezik MT, L'Ecuyer P (2008) Staffing multiskill call centers via linear programming and simulation. *Management Sci.* 54(2):310–323.
- Danaher PJ, Wilson IW, Davis RA (2003) A comparison of online and offline consumer brand loyalty. *Marketing Sci.* 22(4):461–476.
- Danaher PJ, Bonfrer A, Dhar S (2008) The effect of competitive advertising interference on sales for packaged goods. *J. Marketing Res.* 45(2):211–225.
- Deleersnyder B, Geyskens I, Gielens K, Dekimpe MG (2002) How cannibalistic is the Internet channel? A study of the newspaper industry in the United Kingdom and The Netherlands. *Internat. J. Res. Marketing* 19(4):337–348.
- Dube J-P, Hitsch G, Manchanda P (2005) An empirical model of advertising dynamics. *Quant. Marketing Econom.* 3(2):117–144.
- Fader P, Hardie BGS (2000) A note on modelling underreported Poisson counts. *J. Appl. Statist.* 27(8):953–964.
- Fader P, Hardie BGS (2009) Probability models for customer-base analysis. *J. Interactive Marketing* 23(1):61–69.
- Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* 5(2):79–141.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2009) *Bayesian Data Analysis* (Chapman & Hall/CRC Texts in Statistical Science, New York).
- Geykens I, Gielens K, Dekimpe MG (2002) The market valuation of Internet channel additions. *J. Marketing* 66(2):102–119.
- Gijzenberg MJ, van Heerde HJ, Dekimpe MG, Nijs VR (2011) Understanding the role of adstock in advertising decisions. Available at SSRN: <http://ssrn.com/abstract=1905426>.
- Goodhardt GJ, Ehrenberg ASC, Chatfield C (1984) The dirichlet: A comprehensive model of buying behaviour (with discussion). *J. Roy. Statist. Soc., Ser. A* 147(5):621–655.
- Inman JJ, Shankar V, Ferraro R (2004) The roles of channel-category associations and geodemographics in channel patronage decision. *J. Marketing* 68(2):51–71.
- Jeuland AP, Bass FM, Wright GP (1980) A multibrand stochastic model compounding heterogenous Erlang timing and multinomial choice processes. *Oper. Res.* 28(2):255–277.
- Kim S-H, Chan CW, Olivares M, Escobar G (2015) ICU admission control: An empirical study of capacity allocation and its implication on patient outcomes. *Management Sci.* 61(1):19–38.
- Kingstone S (2006) Great expectations: Self-service success can happen. Yankee Group Report, May. Accessed April 2, 2015, http://www.ianbrooks.com/useful-ideas/articles_whitepapers/Great%20expectations.pdf.
- Kleinrock L (1975) *Queueing Systems. Volume 1: Theory* (John Wiley & Sons, Hoboken, NJ).
- Knox GAH (2006) Modelling and managing customers in a multichannel setting. Dissertations available from ProQuest. Paper AAI3225484. <http://repository.upenn.edu/dissertations/AAI3225484>.
- Kumar A, Telang R (2012) Does the Web reduce customer service cost? Empirical evidence from a call center. *Inform. Systems Res.* 23(3, part 1 of 2):721–737.
- Lu Y, Musalem A, Olivares M, Schilkrot A (2013) Measuring the effect of queues on customer purchasing behavior. *Management Sci.* 59(8):1743–1763.
- Mandelbaum A (2006) Call centers: Research bibliography with abstracts. Technion–Israel Institute of Technology. Accessed April 2, 2015, <http://iew3.technion.ac.il/serveng/References/cbib.pdf>.

- McFadden D (1973) Conditional logit analysis of qualitative choice behavior. Zarembka P, ed. *Frontiers of Econometrics* (Academic Press, New York), 105–142.
- Mehrotra V, Fama J (2003) Call center simulation modeling: Methods, challenges, and opportunities. Chick S, Sanchez PJ, Ferrin D, Morrice DJ, eds. *Proc. 2003 Winter Simulation Conf.* (IEEE Press, Piscataway, NJ), 135–143.
- Mehrotra V, Ozluk O, Saltzman R (2010) Intelligent procedures for intraday updating of call center agent schedules. *Production Oper. Management* 19(3):353–367.
- Meuter ML, Bitner MJ, Ostrom AL, Brown SW (2005) Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies. *J. Marketing* 69(2): 61–83.
- Moe WW, Fader PS (2004) Dynamic conversion behavior at e-commerce sites. *Management Sci.* 50(3):326–335.
- Neslin SA, Grewal D, Leghorn R, Shankar V, Teerling ML, Thomas JS, Verhoef PC (2006) Challenges and opportunities in multi-channel customer management. *J. Service Res.* 9(2):95–112.
- Netessine S, Tang C (2009) *Consumer-Driven Demand and Operations Management Models*. International Series in Operations Research and Management Science, Series ed. (Springer, New York).
- Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. *Marketing Sci.* 27(2): 185–204.
- Pierson M, Allon G, Federgruen A (2011) How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing Service Oper. Management* 13(4):489–507.
- Schmittlein DC, Bemmaor AC, Morrison DG (1985) Why does the NBD model work? Robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Sci.* 4(3):255–266.
- Shankar V, Smith AK, Rangaswamy A (2003) The relationship between customer satisfaction and loyalty in online and offline environments. *Internat. J. Res. Marketing* 20(2):153–175.
- Shen H, Huang JZ (2008) Interday forecasting and intraday updating of call center arrivals. *Manufacturing Service Oper. Management* 10(3):391–410.
- Sousa R, Voss CA (2006) Service quality in multi-channel services employing virtual channels. *J. Service Res.* 8(4):356–371.
- Soyer R, Tarimcilar MM (2008) Modeling and analysis of call center arrival data: A Bayesian approach. *Management Sci.* 54(2): 266–278.
- Sun B, Li S (2011) Learning and acting upon customer information: A simulation-based demonstration on service allocations with offshore centers. *J. Marketing Res.* 48(1):72–86.
- Taylor JW (2008) A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Sci.* 54(2):253–265.
- Taylor JW (2012) Density forecasting of intraday call center arrivals using models based on exponential smoothing. *Management Sci.* 58(3):534–549.
- Weinberg J, Brown LD, Stroud JR (2007) Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. *J. Amer. Statist. Assoc.* 102(480):1185–1198.
- Xue M, Hitt LM, Harker PT (2007) Customer efficiency, channel usage, and firm performance in retail banking. *Manufacturing Service Oper. Management* 9(4):535–558.