Learning and Acting Upon Customer Information:

A Simulation-Based Demonstration on Service Allocations with Offshore Centers

Baohong Sun
Tepper School of Business
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA15213
Tel: 412-268-6903
Fax: 412-268-7357
Email: bsun@andrew.cmu.edu

Shibo Li
Kelley School of Business
Indiana University
1309 E. 10th Street
Bloomington, IN 47405
Phone: 812-855-9015
Fax: 812-855-6440
Email: shili@indiana.edu

Baohong Sun is Professor of Marketing at Carnegie Mellon University. Shibo Li is Assistant Professor of Marketing at Indiana University. The authors thank participants at a Marketing Seminar at NYU, Wharton, Columbia, Chicago, Dartmouth, INSEAD, University of Texas at Dallas, University of Toronto, University of California at San Diego, the Biennial MSI Young Scholars Program, and the International Workshop on Customer Relationship Management: Data Mining Meets Marketing for helpful comments. The second author also thanks the Center for International Business Education and Research (CIBER) at Indiana University for awarding a faculty research grant to support this research. The authors also thank Laurence Debo for helpful comments. Both authors contributed equally to the paper and we randomly rotate the order of the authorship.

Learning and Acting Upon Customer Information:
A Simulation-Based Demonstration on Service Allocations with Offshore Centers

Abstract

As service centers become crucial corporate assets for increasing customer relationships and profits, it is imperative to understand customer reactions to service allocations. Using customer call history from a DSL service, the authors empirically investigate how customers' onshore and offshore experience affects service duration and customer retention. They formulate service channel allocation decisions as solutions to a dynamic programming problem in which the firm learns about heterogeneous customer preferences, balances short-term service costs with long-term customer retention, and optimally matches customers with their preferred centers to maximize long-term profit. They demonstrate through simulations that learning enables a firm to make more customized allocations, and acting on long-term customer responses prompts the firm to make proactive decisions that prevent customers from leaving. As a result, the firm can improve customer retention and profit. The proposed framework also mirrors the recent trend of companies seeking solutions that transform customer information into customized and dynamic marketing decisions to improve long-term profit.

Keywords: service outsourcing; channel allocation; customer retention; long-term customer value; customer information management; decision support system; customer-centric marketing; stochastic dynamic optimization

Call centers initially were built to deal with customer inquiries, so their management traditionally has been considered little more than a cost to be saved. This attitude has led to the increasing popularity of outsourcing: More than 3 million agents are employed overseas, likely to increase by 10% per year (McKinsey Quarterly 2005). Most outsourced operations concentrate in the Philippines and India. Early adopters have achieved savings of 40% or more, generally by operating at significant cost scales. However, a recent survey by Purdue University (2004) indicates that despite these significant cost savings, both consumer and business customers report lower satisfaction ratings with outsourced call centers. Outsourcing firms thus have realized that the benefit of driving down costs comes at the cost of alienating customers; in some cases, customer defections even outweigh the potential savings derived from outsourcing (with Purdue University 2004).

Modern call centers handle customer surveys, telemarketing, product inquiries, sales, transactions, promotions, cross-selling, advertising, and postpurchase service by telephone, e-mail, fax, or Web pages. In performing integrated marketing functions, these centers are becoming a preferred customer interaction channel. Approximately 80% of a firm's interactions with its customers occur through call centers, and 92% of customers form their opinions about a firm on the basis of their experience with call centers (Purdue University 2004). The call center industry is among the first to become equipped with the most advanced technology, which offers the capacity to store detailed customer history, retrieve real-time customer information, automatically analyze customer preferences, and instantly respond with highly customized and interactive marketing decisions. For example, the wide adoption of the sophisticated automatic call distributor (ACD), an automated switch designed to route calls, allows managers to monitor the progress and flow of work done by agents, routinely collect information on each agent's call length, analyze a wealth of statistical models about agents in real-time, and automatically route calls (Belt, Richardson and Webster 2000). Most firms also have customer relationship management (CRM) system in place with rich customer information, though few are integrated with ACD systems, so customer information has been largely ignored in service allocation decisions. With high power technology already in place, managers seek analytical decision-making tools that will enable them to integrate rich customer information into the ACD system and generate one-on-one, interactive, dynamic marketing interventions to transform their service centers into revenue growth systems. In short, the meaning of call centers has shifted, from a cost to be saved to a crucial corporate asset, because of their ability to enhance customer relationships and firm profits.

Using customer call history data provided by a DSL service company that operates offshore centers, we develop consumer response models to investigate (1) how customers evaluate the performance of offshore service centers and (2) the relationship among service allocation, service costs, and customer retention. Empirical evidence shows that in general, customers are more likely to leave when they receive service from offshore centers and experience extra-long service duration. However, these sensitivities become less pertinent when the offshore centers handle technical questions. In addition, some customers tend to incur longer service duration, but they also care less about being serviced by offshore centers, especially when they have technical questions.

We next conduct empirically-based simulations and investigate a potential way to use offshore centers better, without significantly jeopardizing customer retention. Instead of proposing cost-ineffective solutions, such as training service agents, we explore the possibility of improving services by using information about customers and acting upon it. In other words, we study how firms can use recent information to learn about customers and allocate customers to their preferred centers, thus continuously improving relationships with customers to maximize their long-term profit contributions. We formulate a firm's service allocation decisions as a matching problem whose solutions derive from a stochastic dynamic control problem with long-term marketing consequences, adaptive learning, and forward planning. Specifically, by allowing service duration and customer retention to be customer-specific and driven by allocation decisions, we let the firm update its knowledge about customer preferences, trade off between short-term service costs and long-term customer reactions, and make optimal allocation decisions that maximize long-term profit. The service allocation problem is formulated as a channel assignment problem in which the firm attempts to improve customer experience and long-term profit contributions by continuously learning about individual customer preferences and then matching customers with their preferred channel in a customized and dynamic fashion.

On the basis of the estimated parameters that characterize the relationship among service allocation, duration, and retention, we conduct simulations with our proposed framework and compare the simulated performance implications with four nested alternative optimization frameworks. It is shown that taking customer preferences into account reduces attrition rate by 21.4%. There will be additional reduction of attrition rate of 4.4%, 14.9%, and 6.9% by adding customer retention, customer lifetime value, and allowing the firm to adaptively learn customer's type. These improvements in customer retention can be translated into 7.0%, .3%, 1.6%, and .8% increase in total profit.

Our study contributes to the understanding of service allocations with offshore centers in several ways. First, our findings provide empirical evidence about how customers evaluate offshore centers and shed new light on the drivers of customer responses to a firm's allocations of service channels. Second, we propose a stochastic dynamic framework with long-term marketing consequences, adaptive learning, forward-looking, and optimization to model a firm's service allocation decisions with offshore centers. Through simulations we demonstrate that the proposed framework may significantly improve the firm's service allocation decisions and achieve better profits and retention rates. Finally, even though our study pertains to service channel allocations, the proposed dynamic framework can be generalized to other marketing mix decisions, such as promotion, cross-selling, and targeted Web advertising.

PRIOR LITERATURE

Marketing offers ample research documenting the relationships among service quality, customer satisfaction, retention, and financial impacts (Anderson and Sullivan 1993; Bolton 1998; Bolton and Drew 1991a, b; Boulding et al. 1993; Boulding, Kalra and Staelin 1999; Kamakura et al. 2005; Li, Sun and Wilcox 2005; Oliver and Swan 1989; Rust and Chung 2006; Rust, Zahorick, and Keiningham 1995; Tse and Wilton 1988). Various models study the link between satisfaction and salesperson incentive schemes (Hauser, Semester and Wernerfelt 1994), the relationship of measured overall service quality and subsequent usage (Bolton and Lemon 1998; Danaher and Rust 1996; Mittal and Kamakura 2001), the explanatory power of customer satisfaction on duration (Bolton 1998), the relationship between customer satisfaction and the firm's productivity level (Anderson, Fornell and Rust 1997), and customer lifetime value (Gupta, Lehmann and Stuart 2004; Reinartz and Kumar 2000, 2003; Rust, Lemon and Zeithaml 2004). Recent "Holy Grail" CRM models claim to determine multiple, personalized marketing interventions over time to manage long-term customer value (Anderson and Salisbury 2003; Bult and Wansbeek 1995; Gonul and Shi 1998; Kamakura et al. 2002; Lewis 2005; Netzer, Lattin and Srinivasan 2009; Rust and Verhoef 2005; Venkatesan and Kumar 2004). However, despite the increasing importance of call centers in shaping customer service experience, no marketing research specifically investigates how customers react to a firm's service channel assignment and studies how the resulting service treatment might affect customer attrition and long-term customer value.

Our proposed framework has the following characteristics. First, with the exception of Gonul and Shi (1998), Lewis (2005), and Rust and Verhoef (2005), most existing work focuses on developing customer response models and discusses implications of the firm's CRM intervention decisions tangentially. We treat the firm as a decision maker and analytically derive the explicit allocation decisions that match customers with the most appropriate centers. Second, most research emphasizes better approaches to model customer heterogeneity, based on demographic variables and pooled historical data, which result in score ranking of customers. We instead propose adaptive learning through continuous interactions, during which customer feedback from the most recent decision execution gets adopted and integrated into the firm's periodic decisions. Third, we treat the firm as a forward-looking decision maker that incorporates the long-term profit implications of customer attrition into its decisions, so future consequences affect the derived optimal decisions.

INDUSTRY BACKGROUND AND DATA DESCRIPTION

The data for our study come from a firm that sells DSL services to both residential and business customers. This firm operates service centers in the United States and globally. For simplification, we treat all service centers within the continental United States as onshore service centers and those outside as offshore service centers. Depending on the speed of the modem, customers pay either $49.95 or $29.95 as a monthly subscription fee to maintain their access to DSL services. Some initial subscriptions require a one-year contract, but customers can terminate the service at any time, with a penalty if the contract is terminated prematurely.

All customers have access to free, 24/7, live customer support. For simplification, we classify customer questions as technical or transactional. Technical questions include software- or hardware-related issues; questions regarding installation, dial-up, user identifications, or passwords; and downed services or network outages. Transactional questions include inquires about billing, e-mail accounts, product news, product services, and registration. When a customer calls, he or she is asked by the automated system to punch in the type of questions before routing. The customer may experience some waiting time before an agent addresses the call, without knowing for which center he or she is waiting. After a call is picked up by an agent, the customer takes a few minutes to describe his or her problem, and then the agent provides solutions. When a call cannot be solved in a timely fashion, the customer may be put on hold while the agent processes the case or sends it to higher-level managers. This scenario occurs more frequently at the offshore centers and for transactional questions, whose frontline agents have less authority to make decisions and escalate more cases to supervisors in their center. In addition, offshore centers are often less-well equipped to solve technical problems.

Because of significant labor costs, the firm calculates service costs primarily on the basis of the labor costs related to the total time that agents remain occupied with a case.[1] Accordingly, the company measures the service duration as the total time of the service encounter—from the time the phone is picked up by an agent to the time the problem is solved. This measure includes time speaking with the customer, as well as time during which the customer is on hold and the agent is processing the customer's request. Multiple calls initiated by the same customer for the same problem usually are routed to the same agent, and the firm's policy states that agents should solve customer problems while customers remain on the phone, with very rare exceptions when agents must perform some task after customers hang up. This measurement is consistent with operation management literature (Gans, Koole and Mandelbaum 2003).

When a customer calls in, the ACD system automatically calculates the average service duration of each agent in handling this type of question and routes the incoming call to the available agent with the lowest estimated service costs. It recognizes the average differential traits of service providers and assigns calls to the one with the lowest average service duration in a myopic way; that is, it ignores customer aspects completely when calculating the service cost. Because this service allocation rule is determined primarily by the estimated service costs, not customer heterogeneity, we term this routing rule "cost-based routing."

[1] Our definition of service duration includes both talking time and (possible) holding time. We do not include waiting time as part of the service duration because from the firm's perspective, only talking time and holding time keep the agent occupied and directly affect service costs. Other than some negligible phone costs, waiting time does not incur labor costs under the ACD system. However, we include waiting time as part of the customer service experience in the retention equation, which we describe subsequently, to take into account its effect on customer attrition. Accordingly, when we run simulations, we consider the different waiting times caused by the proposed service allocation decisions (see Equation 16). Because of the way the company collects data, we cannot separate talking time from holding time, which represents a limitation of our data set, but we do not expect our results to be significantly altered.

Our calibration sample contains the service history of 9,643 calls (calls to disconnect service are not included) initiated by 2,106 randomly selected customers during 52 weeks between January 2003 and December 2003. Our holdout sample contains 1,053 customers who made a total of 4,661 calls. We have access to detailed information about each call, such as the caller's location, time stamps, call reasons, service allocation, call-center agent, call-center manager, and total service duration. In addition, the company randomly selected customers to participate in a satisfaction survey conducted between January and March 2003. Most customers participated in only one satisfaction survey. Because these scores pertain to overall satisfaction during the first three months of our observation period, we treat them as customer summary evaluations of the company's service prior to our observation period. Furthermore, we have customer demographic information, including tenure with the firm, expertise with computers, and number of computers, as well as whether a customer left the firm during the observation period. Finally, we received estimates of marginal service costs, calculated on the basis of the call center agent's wage and other variable costs. The average cost per minute of offshore centers is roughly one-third that of onshore centers.

In Table 1A, we list the definitions and sample statistics of the customer variables. For example, customers paid $99 to terminate their contract prematurely in 1.58% of all observation occasions. Mostly (62%) residential as opposed to business, these customers initiated an average of 6.01 service calls per person, and 90% were technical questions. The average waiting times, divided by four time periods during a day (8–12, 12–16, 16–20, and 20–24), for both centers are approximately 2, 1, 2, and .5 minutes for the onshore centers and 1, .5, 1.5, and .5 minutes for offshore centers. 16% of customers left during the observation period. We choose weeks as our unit of analysis and rate variables, such as prices, accordingly.

In Table 2, we list and compare the allocation, service duration, and retention rates between centers and question types[2]. The current cost-based routing results in 84% of calls being assigned to onshore service centers and 16% handled by offshore centers. Among all questions handled by onshore centers, 11% are transactional and 89% are technical. The split is 3% and 97% for offshore centers. The average service durations are 6.39 minutes for transactional questions and 22.32 minutes for technical questions for onshore centers (cf. 44.20 and 36.28 minutes, respectively, for offshore centers). We note that the difference in technical questions is much lower than that for transactional questions, despite the longer time offshore centers require to solve both types of questions. Between centers, frequent service by offshore service centers leads to higher average customer attrition (17% versus 12% for onshore, $t = 12.6$). Thus, the data suggest that customers prefer onshore centers in terms of retention, though the extent of this onshore preference differs across question types. Although customers are significantly less likely to leave when onshore centers handle their transactional questions (.91 versus .82, $t = 11.3$), the difference in retention rates is much smaller for technical questions (.87 versus .84, $t = 4.6$).

This analysis provides preliminary evidence that though it takes more time for offshore centers to solve both types of questions, the difference for technical questions is much smaller. Taking into account the significantly lower marginal service costs, the offshore centers in our data set have some cost advantages compared with onshore centers for handling technical questions. In the meanwhile, though customers prefer to be serviced by onshore centers, according to their retention rates, they are less sensitive for technical questions handled by offshore centers.

CUSTOMER RESPONSES

[2] We compare the percentages of question types, corresponding service duration, and retention rates across those who made different numbers of calls during the observation period. We find no significant variation in the types of questions or retention rates between frequent and infrequent callers.

We assume that the firm operates j = 1, 2 service centers, with j = 1 representing onshore centers and j = 2 representing offshore centers. At time $t = 1, \ldots, T$, customer $i = 1, \ldots, I$ may call in with question types k = 1 or 2, with k = 1 representing transactional questions and k = 2 representing technical questions.[3] We also assume there are $m = 1, \ldots, M$ segments of customers.[4]

We use three dummy variables, D_{ikt} for k = 0, 1, 2, denote whether customer i calls with question type k at time t, with k = 0 representing the case when customer i does not call. Thus, $D_{i0t} + D_{i1t} + D_{i2t} = 1$. Note that D_{ikt} recognizes call and no-call occasions and is not a decision variable. We use two dummy variables A_{ijt} to denote the firm's allocations decisions, such that $A_{i1t} = 1$ if the question is allocated to an onshore center and $A_{i2t} = 1$ if the question is allocated to an offshore center, with $A_{i1t} + A_{i2t} = 1$. These variables equal 0 when customer i does not call in at time t, that is, when $D_{i0t} = 1$. To simplify our notations, we define a vector $D_{it} = (D_{i0t}, D_{i1t}, D_{i2t})'$ and the vector $A_{it} = (A_{i1t}, A_{i2t})'$.[5]

Service Duration

Intuitively, service duration could be determined by the traits of service centers, question types, and customers. Following Mandelbaum and Schwartz (2002), who show that call duration can be captured best by a log-normal distribution, we assume the log of call duration $\log(DUR_{ijkc}(m))$ for customer i of segment m for all call occasions is given by

$$\begin{aligned} \log(DUR_{ijkc}(m)) = {} & \alpha_0(m) + \alpha_1(m)D_{i2c} + \alpha_2(m)A_{i2c} + \alpha_3(m)A_{i2c}D_{i2c} \\ & + \alpha_4(m)\log(DUR_{ijkc-1}(m)) + \alpha_5(m)NCOMPUTER_i + \xi_{ijkc}(m) \end{aligned} \tag{1}$$

---

[3] For simplicity, we consider two service centers and two types of questions; we ignore differences in service duration among agents within the same center. The proposed approach can be generalized to incorporate multiple service centers, multiple questions, and multiple agents.

[4] Because our research goal is to demonstrate the value of firm learning and acting on customer information, we focus on modeling customer interactions with service centers and assume away possible price endogeneity, capacity adjustments, and extra dynamics introduced by agent learning. These simplified assumptions are reasonable for our research context; we leave the relaxation of these assumptions to further research.

[5] As mentioned earlier, the data were generated by cost-based routing. By ignoring customer preferences, cost-based routing does not take into account customer-specific duration or the center the customer prefers; the allocation depends purely on question type and the average service duration of each center and is a myopic decision. No obvious endogeneity problem exists regarding the estimation of the customer response models. We estimate customer response models for each center and each question type to control for differences between centers and question types.

(2) and $\xi_{ijkc}(m) \sim N(0, \sigma_{\xi}^{2}(m))$,

for all j = 1, 2 and k = 1, 2. We specify the duration equation for each question type and each service center to address the differences in service duration for different centers and question types. The subscript c denotes the counting index of all occasions of calls actually placed.[6] We include the dummy variables D_{i2c} and A_{i2c} to control for the differential service duration across question types and centers. Their coefficients $\alpha_1(m)$ and $\alpha_2(m)$ indicate whether it takes more or less time for service centers to handle technical questions and whether it takes offshore centers more or less time to solve a case. To determine whether the difference in service duration between the centers varies across question types, we include the interaction term $A_{i2c}\, D_{i2c}$, whose coefficient $\alpha_3(m)$ indicates how technical questions modify the difference in the service duration of offshore centers. If $\alpha_2(m) > 0$ and $\alpha_3(m) < 0$, offshore centers are generally slower than onshore centers. However, the difference is smaller when offshore centers handle technical questions. The variable $\log(DUR_{ijkc-1}(m))$ is the log of the total service time it took center j to solve question type k in the prior service occasion. Its coefficient $\alpha_4(m)$ captures the persistence of service duration for customer i. We include $NCOMPUTER_i$, or the number of computers owned by the caller, to note the possibility that customers with more computers may incur longer service times. The coefficient $\alpha_5(m)$ measures the effect of this variable on service duration. We use the vector $\alpha(m)$ to represent all coefficients in Equation 1, and $\xi_{ijkc}(m)$ denotes all other unobserved factors that may affect service duration.

Equation 1 thus defines the service durations for all call occasions. When customers do not call, the duration is 0. For all observation periods $t = 1, \ldots, T$, the expected duration when question type k is allocated to center j is given by

[6] Subscript c represents call occasions. This notation appears only in this equation. As explained subsequently, we define a retention equation for each period t because customers can leave during non-call occasions too.

$$(3)\quad E[DUR_{ijkt}(m)] = \begin{cases} \exp\{E[\log(DUR_{ijkt}(m))] + \frac{1}{2}Var[\log(DUR_{ijkt}(m))]\} & \text{when } D_{i1t} = 1 \text{ or } D_{i2t} = 1, \\ 0 & \text{when } D_{i0t} = 1. \end{cases}$$

Customer Retention

Assume that at each time $t = 1, \ldots, T$, customer i of type m decides whether to stay with the firm. We follow extant marketing literature and allow customer satisfaction, price, presence of competition, switching costs, and variables describing the customer service experience with onshore and offshore centers to drive retention (Bolton 1998). Specifically, the retention decision results from the following function $W_{it}(m)$:

$$(4)\quad \begin{aligned} W_{it}(m) &= \beta_0(m) + \beta_1(m)SAT_i + \beta_2(m)FEE_{it} + \beta_3(m)PROM_{it} + \beta_4(m)COMP_{it} \\ &+ \beta_5(m)\log(TENURE_{it}) + \beta_6(m)PENALTY_{it} + \beta_7(m)A_{i2t} + \beta_8(m)A_{i2t}D_{i2t} \\ &+ \beta_9(m)\sum_{j,k}(D_{ikt}A_{ijt}\log(DUR_{ijkt}((m))) + \beta_{10}(m)(\sum_{j,k}(D_{ikt}A_{ijt}\log(DUR_{ijkt}((m))))^2 + \beta_{11}(m)WAIT_{jt}\, \cdot \\ &+ \beta_{12}(m)\sum_{\tau=1}^{t}\sum_{k=1}^{K}D_{ik\tau} + \beta_{13}(m)FREQ\_OFF_{it} + \beta_{14}(m)ACCUM\_DUR_{it} + \varsigma_{it}(m) \end{aligned}$$

In this equation, SAT_i is the customer's overall satisfaction with the firm, measured at the beginning of the observation period. We expect its coefficient to be positive (Bolton 1998). The variables FEE_{it} and $PROM_{it}$ are fees and promotions specific to customer i at time t. In addition, $COMP_{it}$ is a dummy variable indicating the presence of competitors such as cable service in the geographic area where customer i resides. Switching costs also play an important role in determining customer retention. We include log($TENURE_{it}$) to control for the effect of the length that a customer stays with the firm on customer attrition; the log of this variable allows for its possible nonlinear relationship with customer retention, which may be caused by the one-year contract signed by some customers. To take into account the monetary switching costs, we also include the amount of penalty customers must pay when they terminate the service prematurely.

We are most interested in how variables characterizing customers' experience with onshore and offshore centers affect customer retention. The dummy variable A_{i2t} represents whether the customer is serviced by an offshore center. We include it to capture the effect of being serviced by an offshore center on customer retention. To determine whether having a technical question handled by an offshore center modifies the negative effect on a customer's tendency to leave, we also include the interaction term A_{i2t} D_{i2t}. If $\beta_7(m) < 0$ and $\beta_8(m) > 0$, customers in general are likely to leave when being serviced by offshore centers, but they are less likely to do so when their technical questions are answered by offshore centers.

As the information in Table 2 demonstrates though, service duration varies significantly and may define differential customer service experiences with onshore and offshore centers. We expect the coefficient of $\sum_{j,k}(D_{ikt}A_{ijt}\log(DUR_{ijkt}(m)))$ to be positive, because controlling for everything else (e.g., customers are unhappy that they have to ask questions, expertise, number of computers), customers who need help may appreciate an agent who spends a reasonable amount of time to listen to their explanation of the problem and then provides solutions in a timely fashion.[7] However, customers can quickly become impatient and unhappy when a service call lasts too long, they get put on hold, and/or their cases are escalated to higher-level managers. To account for this effect, we include the squared term of logged service duration and expect its coefficient to be negative. Finally, another component of customer service experience is the time the customer has to wait before his or her call is answered, or $WAIT_{jt}$. We include it to control for its possible negative effect on customer retention (Gans, Koole and Manddelbaum 2003).

We suspect that a service encounter with onshore or offshore centers and service duration may have long-lasting impact on customer retention. The variable $\sum_{\tau=1}^{t}\sum_{k=1}^{K}D_{ik\tau}$ reflects the cumulative number of calls initiated by customer i up to time t. We include this variable to control for the

[7] As Gans, Koole, and Mandelbaum (2003) discuss, very short calls sometimes occur because agents take small "rest breaks" by hanging up on customers.

possibility that customers who face more questions with the service are more likely to leave the company. We also include the recency weighted frequency of being serviced by offshore centers ($FREQ\_OFF_{it}$) and the recency weighted cumulative duration of past service calls prior to time t. Recency weights the duration and frequency, to reflect that events that happen long ago may have less impact on customers, consistent with the idea of using recency, frequency, and monetary value to predict customer responses to direct marketing (Colombo and Jiang 1999; Fader, Hardie, and Lee 2005; Gonul and Shi 1998). The coefficients $\beta_{13}(m)$ and $\beta_{14}(m)$ capture the effects of past experience with offshore centers and service duration on customer retention.

We use the vector $\beta(m)$ to represent all coefficients in customer retention in Equation 4 and $\zeta_{it}(m)$ to represent all the unobservable factors that affect the customer retention decision. Let RET_{it} be a dummy variable indicating whether customer i stays with the firm at time t; then,

(5) $$RET_{it} = \begin{cases} 1, & \text{if } W_{it}(m) \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Let $\overline{W}_{it}(m)$ be the right hand side of Equation 4 without the error term ($\zeta_{it}(m)$). Assuming $\zeta_{it}(m)$ follows an independent, identical extreme value distribution, the probability of customer i staying with the firm at time t can be represented by

(6) $$\Pr_{it}(RET)(m) = \frac{e^{\overline{W}_{it}(m)}}{1 + e^{\overline{W}_{it}(m)}}$$

Heterogeneity and Estimation

We use maximum likelihood to estimate $\alpha(m)$ and $\beta(m)$, which parameterize the effect of service allocation decisions on duration and customer retention. To take into account customer heterogeneity, we adopt a latent class approach, such that segment membership depends on expertise with computers (EXP_i) and type of customer ($RESIDENTIAL_i$ = 1 if residential and 0 if business) (Kamakura and Russell 1989). Denoting the coefficients of these variables as $\gamma(m)$, we define Θ =

$\{\alpha(m), \beta(m), \gamma(m)\}$ for all m as the vector of parameters to be estimated together with the customer-specific duration and retention models.

FRAMEWORK FOR CUSTOMER-CENTRIC ALLOCATION DECISIONS

To improve customer experience and use offshore service centers more effectively, a firm might match each service call with the right center, according to individual customer preferences. We therefore formulate the service allocation decisions of a firm as solutions to a stochastic dynamic programming problem, in which the firm iteratively learns about customer preferences and adapts its allocation decisions to its best knowledge so that it can maximize long-term expected profits.

Adaptive Learning of Customer Heterogeneous Preference

When customer i calls at time t, his or her preference remains unknown or uncertain to the firm. To acknowledge that companies usually conduct segmentation analysis based on demographic variables and know the average probabilities of a customer belonging to a segment, we define the prior belief of customer type $\Pr_{i0}(m)$ as the probability of segment membership, resulting from the latent class approach in the estimation.

In addition to static demographic variables, accrued information obtained by observing customer feedback about the firm's most recent interventions might reveal customer information. There are at least two such information sources: observed prior service duration and observed customer retention. The same customer usually shows a consistent pattern over time in terms of the length of the service duration. For example, retired customers have more time to talk on the phone and likely incur longer service durations. Customer retention observations reveal the customers' reactions to service allocations. For example, if being serviced by an offshore center leads a customer to leave, it implies this customer is very sensitive to offshore centers.

Assuming customers' preferences do not change over time, we let the firm learn about the possibility of a customer belonging to type m, or $\Pr_{it}(m)$ for $m = 1, \ldots, M$. Let $LR_{it}(m = n)$ denote the ratio of the probability of customer i belonging to $m = n$ type, relative to that of $m = 1$ type, perceived by the firm at time t. According to the Bayesian rule of learning, the firm's perceived likelihood ratio of the consumer belonging to type $m = n$ relative to $m = 1$ is given by

$$LR_{it}(m=n) = \frac{\Pr_{it}(m=n)}{\Pr_{it}(m=1)} = LR_{it-1}(m=n)\frac{\Pr(DUR_{ijkt-1} \mid m=n)}{\Pr(DUR_{ijkt-1} \mid m=1)}\frac{\Pr(RET_{it-1} \mid m=n)}{\Pr(RET_{it-1} \mid m=1)} \tag{7}$$

for $m = 1, \ldots, M$. The intuition is as follows: At the beginning of time t, the firm observes new information realized between $t - 1$ and t, namely, the duration DUR_{ijkt-1} and customer retention RET_{it-1}. The firm calculates the probabilities of observed service duration using Equation 1 and observed retention according to Equation 6 for all customer types $m = 1, \ldots, M$. When the joint probability of observing DUR_{ijkt-1} and RET_{it-1}, under the assumption that customer i belongs to segment $m = n$, is greater than that under the assumption that he or she belongs to segment $m = 1$ (or $\frac{\Pr(DUR_{ijkt-1} \mid m=n)}{\Pr(DUR_{ijkt-1} \mid m=1)}\frac{\Pr(RET_{it-1} \mid m=n)}{\Pr(RET_{jt-1} \mid m=1)} > 1$), the likelihood that customer i belongs to $m = n$ group increases. When customers do not call, $\frac{\Pr(DUR_{ijkt-1} \mid m=n)}{\Pr(DUR_{ijkt-1} \mid m=1)}=1$, and the belief gets updated solely on the basis of observed retention. Given the updating rule, for all $m = 1, \ldots, M$, the perceived probability that customer i belongs to type m at time t is given by

$$\Pr_{it}(m=n) = \frac{LR_{it}(m=n)}{\sum_{m=1}^{M} LR_{it}(m)} \tag{8}$$

Dynamic Optimal Allocation Decisions

Following Berger and Nasr (1998), we define a customer's profit contribution to the firm ($PROFIT_{it}$) as the sum of the discounted net contribution, less the firm's cost of serving him or her:

$$\begin{aligned} PROFIT_{it} = \sum_{m=1}^{M} \Pr_{it}(m) RET_{it} \{ & FEE_{it} \\ & - \sum_{k} D_{ikt}[A_{i1t}(C_1 DUR_{i1kt}(m)) + A_{i2t}(C_2 DUR_{i2kt}(m))]\} \end{aligned} \tag{9}$$

The RET_{it} dummy variable indicates whether customer i stays with the firm at time t, and FEE_{it} represents the marginal revenue contributed by customer i at time t. We assume that the fee is paid at the beginning of period t, so the customer remains for the current period and can call to ask questions. In addition, C_1 and C_2 are the unit costs of service for onshore and offshore service centers, respectively.

We define DUR_{it-1}^{H}, RET_{it}^{H}, and $FREQ\_OFF_{it-1}^{H}$ as the entire history of duration, retention, and recency-weighted frequency of being serviced by offshore centers, respectively, observed prior to time t for all j = 1, 2 and k = 1, 2. For simplicity, we use I_{it} = { DUR_{it-1}^{H}, RET_{it}^{H}, $FREQ\_OFF_{it-1}^{H}$ } to denote the most updated information set available to the firm at the beginning of time t. Thus, the expected profit, given the information set I_{it} and allocation decision A_{ijt}, is

$$
\begin{aligned}
&E[PROFIT_{it} \mid I_{it}, A_{it}] = \sum_{m=1}^{M} \Pr_{it}(m)\Pr_{it}(RET)(m)[FEE_{it} \\
(10)\qquad &- \sum_{k=1}^{K} D_{ikt}\{A_{i1t}C_1 E[DUR_{i1kt}(m) \mid I_{it}, A_{it}] + A_{i2t}C_2 E[DUR_{i2kt}(m) \mid I_{it,} A_{it}]\}
\end{aligned}
$$

where $\Pr_{it}(RET)(m)$ is the probability that consumer i of type m will stay with the firm at time t. For periods t + 1 and beyond, $\Pr_{it}(RET)(m)$ relies on Equation 6. Furthermore, $E[DUR_{ijkt}(m) \mid I_{it}, A_{it}]$ is the expected service duration when customer i of type m with question k is allocated to center j at time t, as given by Equation 3, and $\Pr_{it}(m)$ is firm's perceived probability that customer i belongs to segment m, as defined by Equation 8. The expected profit is weighted by the firm's perceived probability of customer i belonging to type m at time t, or $\Pr_{it}(m)$.

In practice, the allocation of services may not be driven solely by expected profit for many reasons. For example, political and ethical considerations may motivate the firm to assign more cases to onshore centers. To address this possibility, we (the researchers) assume that the firm makes allocation decisions between onshore and offshore centers according to a more general function U_{it}:

$$
(11)\qquad U_{it} = \sum_{j} A_{ijt}\lambda_{0j} + \lambda_1 E[PROFIT_{it} \mid I_{it}, A_{it}] + \sum_{j,k}(D_{ikt}A_{ijt}\tau_{ijkt})
$$

Scalar λ_{01} captures the firm's intrinsic preference to allocate a service call to onshore centers due to factors that are unobserved to the researchers (utilization of capacity is approximated by waiting time as we will explain below), λ_1 measures the importance of expected profit in determining the firm's allocation decisions (relative to the error term). Equation 11 describes a more realistic situation that nests the special case when the firm's allocation decision is driven solely by financial considerations. τ_{ijkt} denotes all the unobserved random factors that affect the firm's allocation decisions, and $\sum_{j,k}(D_{ikt}A_{ijt}\tau_{ijkt})$ is 0 when customer i does not call at time t (e.g., when D_{i0t} = 1 and A_{ijt} = 0). It is τ_{ijkt} when customer i calls in with question of type k and gets allocated to center j. Assuming that τ_{ijkt} has an I.I.D. extreme value distribution, we obtain a binary logit model that approximates the firm's allocation decisions. We also define $\overline{U}_{it}$ as the deterministic part of U_{it}. When a customer calls, the probability of the firm making decision A_{ijt} is given by

(12)
$$\Pr(A_{ijt} \mid D_{ikt}) = \frac{\exp(\overline{U}_{it}(A_{ijt}))}{\sum_j \exp(\overline{U}_{it}(A_{ijt}))}.$$

The values of λ_{0j} and λ_1 are estimated from the binary logit model independent from the customer response models.

To take into account future marketing consequences, the firm must trade off the current costs of service (service costs) and future customer retention (marketing consequences) to maximize its long-term profit (long-term customer value). The firm's decision process can be parsimoniously formulated as the solutions to a stochastic dynamic programming problem, in which the firm makes allocation decisions to maximize its long-term objective function obtained from each customer i:

(13)
$$\underset{A_{ijt}}{Max}\, E\left\{\sum_{\tau=t}^{\infty}\delta^{\tau-1}U_{i\tau}\right\},$$

where $0 < \delta < 1$ is the discount factor that reflects that current utility is preferred to future utility (Erdem and Keane 1996; Erdem, Imai and Keane 2003; Krishna 1994a, 1994b; Sun 2005; Sun, Li and Zhou 2006). The optimal allocation decision is thus the solution to the Bellman equation:

(14)
$$V_{it}(I_{it}, D_{it}) = \max_{A_{ijt}}\{U_{it} + \delta E[V_{it+1}(I_{it+1}, D_{it+1})]\}$$

where $E[V_{it+1}(I_{it+1}, D_{it+1})]$ is the expected optimal utility beginning from time t + 1.

To solve the dynamic programming problem at time t, the firm must know whether customer i will call in with question type k from t + 1 forward to calculate expected future utilities. Because the type of current calls likely depends on the types of calls in the past, we adopt a first-order Markov chain with three states to approximate the probabilities of incoming question types. Given a call type from the last period $D_{ikt\text{-}1}$, for $k = 0$, 1, and 2, with the probability $\rho_1(D_{ikt\text{-}1})$, the customer will call with question type k = 1, the probability $\rho_2(D_{ikt\text{-}1})$ is that he or she will call with question type k = 2, and probability 1 - $\rho_1(D_{ikt\text{-}1})$ - $\rho_2(D_{ikt\text{-}1})$ is the customer does not call. The Bellman equation becomes:

$$\begin{aligned}(15)\qquad V_{it}(I_{it}, D_{it}) = \max_{A_{ijt}} U_{it} + \delta E[\rho_1(D_{ikt-1})V_{it+1}(I_{it+1}, D_{i1t+1} = 1) \\ + \rho_2(D_{ikt-1})V_{it+1}(I_{it+1}, D_{i2t+1} = 1)] + (1 - \rho_1(D_{ikt-1}) - \rho_2(D_{ikt-1}))V_{it+1}(I_{it+1}, D_{i0t+1} = 1)\end{aligned},$$

where $V_{it+1}(I_{it+1}, D_{ikt+1} = 1)$ is the optimal value function when customer i calls in with question type k at time t + 1. Following Hendel and Nevo (2006) and Sun (2005), we approximate the values of $\rho_1(D_{ikt\text{-}1})$ and $\rho_2(D_{ikt\text{-}1})$ using sample call-in frequencies for k = 0, k = 1, and k = 2 types of questions (the sample statistics are reported in Table 1B).

Similarly, when making allocation decisions, the firm must predict waiting times and its possible effect on customer retention. Due to data unavailability, however, we do not explicitly model the firm’s capacity constraint in the proposed framework. Instead, we approximate the waiting time as a proportional function of allocation probability. That is, we assume the effective arrival rate is proportional to the allocation probability, and the expected waiting time is proportional to the effective arrival rate around the realized effective arrival rate. Because we observe the average waiting times of the cost-based routing for four time periods during the day for both onshore and offshore centers, we approximate the waiting time as

(16) $$WAIT'_{jt} = \frac{\sum_{i=1}^{I} \Pr(A'_{ijt})}{\sum_{i=1}^{I} A_{ijt}} WAIT_{jt},$$

where $\sum_{i=1}^{I} A_{ijt}$ and $WAIT_{jt}$ are the observed total number of calls allocated to center *j* and the average waiting time of center *j* during the period when customer *i* calls, respectively. $\Pr(A'_{ijt})$ is the percentage of calls allocated to center *j*, suggested by the proposed allocation A'_{ijt}. Assuming the number of callers in the queue is proportional to the allocation probability, Equation 16 adjusts the waiting time upward (downward) when more (less) customers are allocated to center *j*, as suggested by the proposed allocation decisions.

Given the first-order Markov model of customer call-in, the log-normal model of the service duration, and the assumption that the number of agents stays the same at both on-shore and off-shore centers without call abandonment, retrials and returns, the expected waiting time (or the capacity utilization rate) is approximated. Although not a comprehensive model for waiting time, our proposed approach is consistent with the M/M/N model (Gans, Koole and Mandelbaum 2003), a common modeling approach used by both OM academics and practitioners to approximate the stationary call center system performance of short time intervals (like half an hour) when the arrival rate is sufficiently smaller than the departure rate—a reasonable assumption given the long average service duration in our data. Our modeling of waiting time clearly is a simplification. Readers who are interested in comprehensive waiting time models can read Gans, Koole and Mandelbaum (2003) and Mandelbaum (2002).

To solve the dynamic program problem, we undertake two iterative steps: The firm continuously learns about each individual customer by analyzing customer information according to the customer's revealed reactions to the firm's most recent interactions, and then it adapts its decisions according to its most recent knowledge about each customer. During these integrated and iterative processes, updated knowledge continuously adjusts the firm's decisions, and the resulting customer reactions again inform the learning process. Thus, learning and decision making are integrated, which we refer to as "adaptive learning."[8] Adaptive learning enables the firm to follow customers and improve the accuracy of its knowledge about each individual customer. Because the state variables are continuous, we face the problem of a large state space and therefore adopt the interpolation method developed by Keane and Wolpin (1994) to calculate the value functions for a few state space points (the Web Appendix provides more details).

The solution results in a sequence of intertemporally related, optimal allocation decisions for all calls initiated by each customer. The proposed allocation decisions should be customized, because adaptive learning enables the firm to improve its knowledge about each individual customer and allocate calls according to its best knowledge about that customer's preferences. The allocation decisions should also be proactive, because taking into account the long-term consequences of alienating customers, the firm can sacrifice short-term costs by allocating a customer to his or her desired center to prevent defection and improve long-term profits.

EMPIRICAL RESULTS

In this section, we first discuss the parameter estimates Θ based on the calibration sample that characterize the relationship of the firm's allocation decision, service duration, and customer retention and compare various customer response models using both the calibration and holdout samples. On the basis of these estimated parameters, we conduct simulations using our proposed framework to derive the sequence of optimal allocation decisions A_{ijt} based on the calibration sample. We investigate whether and how the proposed approach offers improvements over four nested alternative frameworks.

Estimation

[8] Conjoint analysis research adopts similar ideas to reveal consumer preferences (Toubia et al. 2003).

Model Comparison. To determine whether our proposed model describes customer responses better than do alternative model specifications, we estimate three benchmark customer response models (Models 1–3) with and without the variables that characterize customer experience with onshore and offshore service centers (i.e., service duration and waiting time, service by offshore centers, and interaction term that indicates whether the offshore centers handle technical questions). In Table 3, we report the model-fitting statistics of the three estimated benchmark models and the proposed customer response model. For both the calibration and the holdout samples, our proposed customer response model explains customer retention and service duration better than the three benchmark models, which supports the need to incorporate our proposed variables. Furthermore, we compare the simulated frequency of offshore allocations, average call durations, and probabilities of retention with those from the sample in Table 4. The fit is good on all dimensions, which indicates that the proposed consumer model with two segments approximates the data quite well.

Coefficients. In Table 5, we report the estimation results. In the duration equation, the positive and significant coefficients of D_{i2c} for both segments show that it takes more time to solve technical questions. Similarly, the positive coefficients of A_{i2c} show that it generally takes longer for offshore centers to solve a case. However, when offshore centers handle technical questions, the service durations are significantly shorter than those associated with transactional questions handled by offshore centers. In addition, the positive coefficients of the service duration of the last service call suggest that if the last call initiated by customer i lasts longer, the current call may also last longer. Customers with more computers also are more likely to ask more questions during each call.

The retention equation reveals that overall customer satisfaction makes customers more likely to stay, as do lower prices, promotions, lack of competition, and a long history with the firm. Not surprisingly, a price penalty prevents customers from leaving. When serviced by offshore centers, both customer segments are more likely to leave, consistent with the finding of the Purdue University study.

However, this negative impact is mitigated by technical questions being handled by offshore centers for customers in the second segment, who are less sensitive to the service of technical questions by offshore centers. The log-duration shows a positive effect on retention, such that customers appreciate it when the service agent spends sufficient time to address their questions. However, the negative coefficient of the squared duration indicates that customers react negatively to extra-long service durations. In other words, customers are more likely to leave when their questions are not addressed in a timely fashion. As we expected, both segments react negatively to longer waiting times. The greater the total number of questions, the less likely customers are to stay, which is consistent with the intuition that on average, customers who face more problems are more likely to leave the company. The recency-weighted frequency of being serviced by offshore centers decreases the chance that both segments stay; the recency-weighted total service duration has negative effects on retention, such that customers appear to react negatively to the high total amount of time they incur to ensure the function of their service.

We estimate that 66.3% of the customers appear in the first segment and 33.7% in the second segment. The coefficients of demographic variables in the segment membership suggest that customers with more expertise with computers and business customers are more likely to belong to the first segment, which may explain why they are more sensitive to service quality but not to price compared with those in the second segment. Customers in the two segments differ in their sensitivities to all the variables in the duration and retention equations. However, the most important differentiators are, in the duration equation, the intercept and, in the retention equation, the intercept, coefficients of being serviced by offshore centers, coefficients of the log of the squared duration, and interaction between offshore and technical questions. All else being equal, customers in the first segment experience shorter service durations (i.e., constant term in the duration equation) and are more likely to stay (i.e., constant term in the retention equation). However, they react much more negatively to being serviced by offshore centers and are more sensitive to service times that last too long. In contrast, those in the second segment seem to incur much longer service durations, are less likely to stay, are less sensitive to being serviced by offshore centers, and are slightly more tolerant of extra-long service encounters. Most important, these customers exhibit less sensitivity to whether their technical questions are handled by offshore centers.

In summary, the estimation results reveal that customers have differential sensitivities to onshore and offshore centers. To improve customer experiences with the company and use offshore service centers more effectively, the firm should match each service call with the right center, according to individual customer preferences and the comparative advantage of the onshore and offshore service centers. However, examining the latent class segmentation, it is not intuitively clear which segment should be allocated to offshore centers. For example, even though customers in the first segment are more likely to stay, they are also more sensitive to being serviced by offshore centers. Thus, the firm should balance all factors that affect service duration and customer retention when deciding on service allocation. We next demonstrate how firm learning and acting on that knowledge can yield customized and proactive allocation decisions and improve profit.

Simulation

We conduct simulations to derive optimal allocation decisions using our proposed framework. Customer service duration and retention are simulated using calibration sample and may change with alternative service allocation decisions. We hope to demonstrate the effectiveness of adaptive learning (Figure 1), show how our proposed allocations are tailored to the firm’s updated information about customer preferences (Figure 2, Panels A and B), establish how service allocation decisions may be driven by long-term marketing consequences (Figure 2, Panel C), and compare the improvement for the proposed model with four nested optimization frameworks (Table 7).

On the basis of the estimated parameters (Θ), the observed call history (D_{ikt}), covariates in the customer response model that are exogenously given, customer demographic variables, importance of firm's financial consideration (λ),[9] and probability of future arrival rate (ρ), we simulate optimal allocation decisions (A^{*}_{ijt}). To take customer aspects into account, we incorporate customer-specific predictions of service duration and retention, as specified in Equations 3 and 6. To add adaptive learning, we set the initial probabilities $\Pr_{i0}(m)$ to be the same as those derived from the latent class estimates and update $\Pr_{it}(m)$ periodically according to Equations 7 and 8, using each customer's most recent information. To add the forward-looking component, we follow the convention and set δ to .995 (Erdem and Keane 1996), then obtain the optimization by solving the Bellman equation.

Learning about Customer Type. In Figure 1, we demonstrate the progress of adaptive learning by dividing the whole observation period equally into three stages and comparing the probabilities of the segment 2 customers ($\Pr_{it}(m = 2)$), which the firm obtains at the end of each stage. The firm perceives customers as relatively the same during the first stage, because the latent class approach results in average segment memberships that are the same across customers with the same demographic variables. As adaptive learning continues, uncertainty falls significantly, and distributions start to show two modes in the second stage. At the end of the observation period, almost every customer can be categorized as either a segment 1 or segment 2 customer. Approximately 34% are classified to the second segment at the end of the observation period, which implies that adaptive learning enables the

---

[9] We estimate the values of λ_{0j} and λ_1 using the observed data by applying a binary logit model defined by Equation 12, with expected profit being replaced by $-\{\sum_k D_{ikt}[A_{i1t}(C_1\overline{DUR}_{1kt})+A_{i2t}(C_2\overline{DUR}_{2kt})\}$. The intercept of firm's objective function is estimated at 8.94 with standard deviation of 1.91 (λ_{02} is normalized to 0 for identification purpose), indicating that the firm has a greater tendency to route questions to onshore centers (factors that are different from capacity constraint and are unobservable to researchers). The coefficient of the financial consideration is 14.07 with standard deviation of .002, implying that financial consideration (i.e. cost) plays very important role in determining allocation decisions. Since we do not explicitly model capacity constraint and the approximation of waiting time is subject to assumptions in the proposed framework, the values of λ_{0j} and λ_1 should be interpreted with caution.

firm to use the information about each customer's most recent interaction to pinpoint segment membership.

Properties of Proposed Solution. Using our proposed framework, we obtain a sequence of optimal allocation decisions for all calls initiated by each customer. To describe the allocation decisions succinctly, we present some summary statistics of the proposed allocations in Table 6 and compare them with observed cost-based allocation decisions. Our proposed approach increases the case assignments to offshore centers from 16% to 19%. Even though this increase is marginally noticeable, the composition of question types and customer types change significantly. Among the technical questions, 20% get assigned to offshore centers, an 18% increase from the 17% assigned by the cost-based method. In addition, 45% of customers from segment 2 are allocated to offshore centers, a 32% increase from the 34% observed in the sample. Because they are less sensitive to being serviced by offshore centers, more segment 2 customers get allocated to offshore centers. The average service duration for both centers decreases significantly, and the customer retention rate increases. Because this model considers waiting time and its negative effect on customer retention, the proposed allocations do not incur greater waiting times.

In Figure 2, Panel A, we depict the average probabilities of assigning a customer to an offshore center ($\Pr(A_{i2t}) = \dfrac{\exp(\overline{U}_{it}(A_{i2t}))}{\sum_j \exp(\overline{U}_{it}(A_{ijt}))}$) and the average probabilities that customers are perceived to be segment 2 customers ($\Pr_{it}(m = 2)$). Because this function reflects the relationship between decision variable and state variable, we term it the allocation function. The proposed solution shows that on average, the higher the perceived probabilities of belonging to the second segment, the higher are the probabilities of being routed to offshore centers, because these customers are less sensitive to longer service times and not as sensitive to being serviced by offshore centers. Thus, empowered by adaptive learning, the firm's allocation decisions are tailored to the firm's most updated knowledge about each individual customer. This method warrants a better match between the service center and individual customer.

However, the relationship likely will be modified by exogenous variables such as question type. In Figure 2, Panel B, we therefore compare the allocation functions between transactional and technical questions. For the same perceived likelihood of belonging to the second segment, it is more likely for the firm to allocate technical questions to offshore centers, consistent with our observation that it is less costly for offshore centers to handle technical questions and that customers are less likely to leave if their technical questions are handled by offshore centers. That is, the firm's allocation decisions recognize the comparative advantages of offshore centers for technical questions.

In Figure 2, Panel C, we compare the allocation function of the 16% of customers who left at the end of our observation period and those who stayed under the cost-based routing system. The proposed solution sacrifices noticeable service costs by allocating customers who are most likely to leave to onshore centers to prevent them from leaving. This trend demonstrates the proactive nature of the allocation decisions enabled by the forward-looking and optimization components of the proposed framework.

Alternative Optimization Frameworks. Our proposed framework enhances cost-minimization routing in at least four important components: the human aspects represented by customer-specific service duration, customer retention, adaptive learning, and long-term customer value. We are interested in comparing the performance of our proposed framework with four nested optimization frameworks and investigating how adding each of these components helps the firm improve retention rate and profit. To make the simulation results comparable, we assume the firm maximizes profit (or $\lambda_0 = 0$ and $\lambda_1 = 1$) in all the five competing frameworks. The first alternative framework (Alternative 1, cost-minimization framework without considering customers) is very similar to cost-based routing that generates our data: when determining the allocation of incoming calls, the firm allocates calls to the center with lower average historical service durations, which is calculated as $\sum_k D_{ikt}[A_{i1t}(C_1\overline{DUR}_{1kt}) + A_{i2t}(C_2\overline{DUR}_{2kt})$, the product of the marginal cost (C_j) and the average service duration of center j prior to time t for handling question type k ($\overline{DUR}_{jkt}$). It only recognizes the average differential traits of onshore and offshore centers. Customer aspects are completely ignored. In the second alternative framework (Alternative 2, cost-minimization framework with customer-specific duration), the cost is given by $\sum_{m=1}^{M}\sum_{k=1}^{K}\Pr_{it}(m)D_{ikt}[A_{i1t}(C_1 DUR_{ijkt}(m)) + A_{i2t}(C_2 DUR_{ijkt}(m))]$. Unlike Alternative 1 that calculates service costs based on $\overline{DUR}_{jkt}$, this framework recognizes customers by predicting service costs according to the service duration predicted by customer response Equation 1. The third one (Alternative 3) adds customer retention to the second framework, such that the firm takes customer retention into account but remains myopic. In the fourth framework (Alternative 4), we add the forward-looking customer lifetime value but not continuous learning to the third framework. Finally, the fifth framework also includes adaptive learning; it represents our proposed framework.

We compare performance improvements across the five alternative frameworks in Table 7; every component contributes to improve retention rates and profit. Specifically, compared with cost-minimization routing which completely ignores customer preferences (Alternative 1), cost-minimization considering customer-specific duration (Alternative 2) reduces attrition rate by 21.4%. There will be additional reduction of customer attrition of 4.4% by taking into account customer retention (Alternative 3), an additional 14.9% by adding customer lifetime value (Alternative 4), and an additional 6.9% by further allowing the firm to adaptively learn customer's type (Proposed framework). These numbers correspond to 3.7%, .8%, 2.6%, and 1.2% of improvement in retention rate which can be translated into 7.0%, .3%, 1.6%, and .8% increase in profit over the cost minimization without considering customers.[10] Thus, the more customized and proactive decisions that result from our proposed framework help firms increase customer retention and improve long-term customer value or profit.

Improvement over Cost-Minimization Routing. As indicated by Table 6, attrition rates are approximately 8% with the proposed allocations, a 50% reduction of the attrition rate compared with Alternative framework 1. To investigate how sensitive the improvement is to the mix of customer types, we increase the percentages of business customers from 0% to 100% and report the customer retention rate at the end of observation period (Figure 3, Panel A) and the total profit accumulated over the observation period (Figure 3, Panel B) resulting from our proposed framework. When the percentage of business customers is relatively low, the higher the percentage of business customers, the more effective the proposed approach is in retaining customers. This is because the proposed allocation decisions recognize that business customers are more sensitive to being serviced by offshore centers. Thus, the firm fine-tunes the allocation to retain these customers in the long run. However, improved retention means higher service costs. Profit improvement starts to decrease at the peak of about 30% of business customers.

Similarly, we increase the percentage of transactional questions from 0% to 100%. Customer retention does not change with the mix of question type, because the composition of questions asked does not seem to significantly affect retention (Table 2). However, since technical questions take more time to solve, the saved costs that result from allocating technical questions to offshore centers increase profit. When the percentage of transactional questions increases beyond a threshold (greater than 50% in this context), the profit improvement thereafter remain almost constant. These results demonstrate

[10] We thank the area editor for this suggestion. We cannot directly compare the performance of our proposed framework with that of the observed cost-based routing. We need to have the same λs in the firm's objective function to make the simulation results comparable. Thus, we assume $\lambda_0 = 0$ and $\lambda_1 = 1$ in Equation 11 when preparing Table 7 and Figure 3. This will not significantly affect the comparison results because the comparison is based on the proposed model of the world. We also repeat the same exercise assuming $\lambda_0 = 8.94$ and $\lambda_1 = 14.07$, the values estimated using the observed data. The percentages of reduction in attrition rates are 20.5%, 4.3%, 14.0%, and 6.5%, which shows similar improvement when additional components are considered.

that the improved customer retention and profit resulting from the proposed approach changes with the nature of the customer base and the composition of questions. However, we suggest caution in generalizing these results, because the direction and magnitude depend on other parameters that are specific to the context.

CONCLUSIONS, LIMITATIONS, AND FURTHER RESEARCH

The role of call centers has shifted from a cost to be saved to a preferred and prevalent channel to handle integrated marketing functions, which makes it an increasingly important corporate strategic asset. When this important corporate asset rests in the hands of a third party, outsourcing firms face the challenge of dissatisfied customers and high customer attrition. The call center industry provides an excellent example of the vast possibilities of transforming customer information into service excellence and revenue growth opportunities.

Using panel data pertaining to service allocations, we first provide empirical evidence about how service duration and customer retention may be affected by the firm’s service allocation decisions. Because of their significantly lower service costs per minute, the offshore centers we study have some comparative advantages over onshore centers when it comes to technical questions. Customers have heterogeneous sensitivities to service duration and allocations. Some customers tend to incur longer service durations, but these same customers also care less about being serviced by offshore centers, especially when they have technical questions.

We formulate service allocation decisions as a matching problem in which the firm recognizes customer-specific service costs and the long-term marketing consequences, learns about customers’ heterogeneous preferences, balances the trade-offs between short-term service costs and long-term customer reactions, and makes optimal allocation decisions that best match customer preferences and maximize long-term profit. On the basis of the estimated parameters, we conduct simulations and apply our proposed framework to derive the optimal call allocation decisions and demonstrate that adaptive learning allows the firm to improve its allocation decisions by matching customers with their preferred service centers. Forward-looking and optimization allow the firm to make proactive decisions to act on its knowledge about customers and long-term marketing consequences. The simulation results show that by recognizing customer aspects, take into account customer retention and customer lifetime value, our proposed model significantly reduces attrition rate and also improve profit. In other words, firms can be better off by continuously managing and acting on customer information.

Our results provide the first empirical evidence about understanding customer reactions to a firm's service channel allocation decisions. Through a simulation-based demonstration on a DSL firm's service channel allocations, our proposed learning and optimization solutions (or simplified heuristics) provide a computational algorithm for firms to integrate with their customer database and operating systems and automate call allocations. Learning about customers in a continuous fashion on the basis of their feedback to the firm's most recent decisions, the firm can not only follow the footsteps of a customer but also adapt its decisions to customer preferences. This approach aligns better with the recent technology-enabled trends toward one-on-one, interactive, and real-time marketing decision making. It further meets the demands of various industries that seek analytical decision-making tools to analyze their customer databases and support their day-to-day marketing decisions.

However, we also acknowledge that our study is limited and could be expanded in several ways. First, due to our data constraints, we make some simplified assumptions about the operations management of service allocation decisions, such as queuing, waiting time, abandonment, and retrials. We also cannot separate active talking time from holding time. Further research could examine how these variables affect customer retention. Second, this study is mainly based on simulations. Future research can use field experiments to further test and generalize the results. Third, we assume the customer preferences are static. Research could develop more sophisticated learning routines to allow for dynamic changes in customer preference. Different approaches, such as hierarchical Bayes, could be integrated into the proposed framework to learn more about customer heterogeneity (Allenby, Arora, and Ginter 1998; Allenby and Rossi 1999; Rossi, McCulloch, and Allenby 1996). Fourth, possible price and capacity endogeneity issues could be investigated and capacity constraint may be explicitly modeled. For example, firms usually increase prices when service quality improves. Firms also may adjust their capacity as a result of improvements in efficiency. Lastly, agents at call centers can learn and become more efficient over time and there may be call center heterogeneity on capacity and equipment (i.e. offshore centers may be less well equipped to solve technical problems), so additional research could account for agents' learning curve or the call center heterogeneity.

References

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (3), 384-389.

——— and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (1/2), 57-78.

Anderson, Eugene, Claes Fornell, and Roland T. Rust (1997), "Customer Satisfaction, Productivity, and Profitability: Differences Between Goods and Services," *Marketing Science*, 16(2), 129-145.

Anderson, Eugene and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 12(2), 125-143.

——— and Linda Court Salisbury (2003), "The Formation of Market-Level Expectations and its Covariates," *Journal of Consumer Research*, 30(1), 115-124.

Belt, V., R. Richardson, and J. Webster (2000), "Women Work in the New Economy: The Case of Telephone Call Centers," *Information, Communication and Society*, 3(3)**,** 367-385.

Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12 (1), 17-30.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17(1), 45-65.

——— and James H. Drew (1991a), "A Multistage Model of Customers' Assessments of Service Quality and Value," *Journal of Consumer Research*, 17 (March), 375-384.

——— and ——— (1991b), "A Longitudinal Analysis of the Impact of Service Changes on Customer Attitudes," *Journal of Marketing*, 55 (January), 1-9.

——— and Katherine N. Lemon (1998), "A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction," Marketing Science Institute Working Paper, 98-106, Marketing Science Institute, Cambridge, MA.

Boulding, William, Ajay Kalra, and Richard Staelin (1999), "The Quality Double Whammy," *Marketing Science*, 18(4), 463-484.

———, ———, ———, and Valerie A. Zeithaml (1993), "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions," *Journal of Marketing Research*, 30 (February), 7-27.

Bult, Jan-Roelf and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science,* 14 (4), 378-394.

Colombo, Richard, and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, 13 (3), 2-12.

Danaher, Peter J. and Roland T. Rust (1996), "Indirect Financial Benefits from Service Duality Quality," *Management Journal*, 3 (2), 63-75.

Erdem, Tülin and Michael P. Keane (1996), "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1-20.

———, Susumu Imai, and Michael P. Keane (2003), "Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, 1 (1), 5-64.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (November), 415-430.

———, G. Koole, and A. Mandelbaum (2003), "Telephone Call Centers: Tutorial, Review and Research Prospects," *Manufacturing and Service Operations Management*, 5 (2), 5:79–141.

Gönül, Fusun and Mengze Shi (1998), "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science,* 44(9), 1249-1262.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41 (1), 7-18.

Hauser, John R., Duncan I. Semester, and Birger Wernerfelt (1994), "Customer Satisfaction Incentives," *Marketing Science*, 13(4), 327-350.

Hendel, Igal, and Aviv Nevo (2006), "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74 (6), 1637-1673.

Kamakura, Wagner, Carl F. Mela, Asim Ansari, Anand Bodapati, Peter Fader, Raghuram Iyengar, Prasad Naik, Scott Neslin, Baohong Sun, Peter C. Verhoef, Michel Wedel and Ron Wilcox (2005), "Choice Models and Customer Relationship Management" *Marketing Letters*, 16 (3), 279-291.

———, Vikas Mittal, Fernando de Rosa, and Jose Afonso Mazzon (2002), "Assessing the Service-Profit Chain," *Marketing Science*, 21(3), 294-317.

———, and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation," *Journal of Marketing Research,* 26 (November), 379-90.

Keane, Michael P., and Kenneth I. Wolpin (1994), "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76 (4), 648-72.

Krishna, Aradhna (1994a), "The Effect of Deal Knowledge on Consumer Purchase Behavior," *Journal of Marketing Research,* 31 (1), 76-91.

———, (1994b), "The Impact of Dealing Patterns on Purchase Behavior," *Marketing Science,* 13 (4), 351-373.

Lewis, Michael (2005), "A Dynamic Programming Approach to Customer Relationship Pricing," *Management Science*, 51 (6), 981-994.

Li, Shibo, Baohong Sun, and Ronald T. Wilcox (2005), "Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services," *Journal of Marketing Research,* XLII (May), 233-239.

Mandelbaum, Avishai (2002), "Call Centers (Centres): Research Bibliography with Abstracts," Version 3. Available at ie.technion.ac.il/serving/References/ccbib.pdf.

———, R. Schwartz (2002), "Simulation Experiments with M/G/100 queues in the Halfin-Whitt (Q.E.D.) Regime," Technical Report, Technion, Haifa, Israel. Downloadable from ie.technion.ac.il/serveng/References/references.html.

McKinsey Quarterly (2005), "Getting More from Call Centers," April, Keith A. Gilson and Deepak K. Khandelwal.

Mittal, Vikas and Wagner A. Kamakura (2001), "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing Research*, 38 (February), 131-42.

Netzer, Oded, James Lattin, and V. Srinivasan (2009), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27 (2), 185-204.

Offshore Digest (2005), "Offshore Outsourcing Reports."

Oliver, Richard L. and John E. Swann (1989), "Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach," *Journal of Marketing*, 53 (April), 21-35.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321-340.

Purdue University, Center for Customer-Driven Quality (2004), "Offshore Company Call Centers a Concern to U.S. Consumers." West Lafayette, IN: Purdue University.

Reinartz, Werner and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractural Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17-35.

———, and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67(1), 77-99.

Rust, Roland T. and Tuck Siong Chung (2006), "Marketing Models of Service and Relationships," *Marketing Science*. 25 (6), 560-580.

———, Anthony J. Zhorick, and Timothy L. Keiningham (1995), "Returns on Quality (ROQ): Making Service Quality Financially Accountable," *Journal of Marketing*, 59 (April), 58-70.

———, Katherine N. Lemon, and Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68 (January), 109-127

———, and Peter C. Verhoef (2005), "Optimizing the Marketing Interventions Mix in Intermediate-Term CRM," *Marketing Science*,24 (3), 477- 489.

Sun, Baohong (2005), "Promotion Effect with Endogenized Consumption," *Marketing Science*, 24(3)**,** 430-443.

———, Shibo Li, and Catherine Zhou (2006), "'Adaptive' Learning and 'Proactive' Customer Relationship Management," *Journal of Interactive Marketing*, 20 (3-4), 82-96.

Tse, David K. and Peter C. Wilton (1988), "Models of Consumer Satisfaction Formation: An Extension," *Journal of Marketing Research*, 25 (May), 204-212.

Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation," *Marketing Science*, 22(3), 273-303.

Venkatesan, Rajikumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68 (4), 106-125.

Table 1A. Variable Definitions and Sample Statistics

Variable	Definition	Mean (SD)
TENURE	Number of months with the service provider since first purchase.	20.29(12.37)
PRICE	Price of the product plan.	$43.91(7.80)
PROM	One-time price promotion for the product.	$4.13(15.97)
COMPET	Dummy variable indicating the presence of competitive service such as cable.	.17(.37)
PENALTY	Penalty fee for terminating a contract prematurely.	$99.00(.00)
NCOMPUTER	Number of computers owned by the caller.	1.63(.77)
EXP	Caller expertise self-rating: 1=extremely inexperienced/novice; 5=extremely experienced/expert;	3.11(1.02)
RESIDENTIAL	Whether the caller is a residential customer.	.62(.49)
NCALLS ()	Total cumulative number of calls.	6.01(18.15)
TECHNICAL	Whether the call is about a technical question.	.90(.30)
FREQ_OFF	The recency-weighted frequency of being serviced by offshore centers.	.30(.40)
SAT	Overall satisfaction rating of the overall service satisfaction quality of the firm.	3.40(1.29)
RET	Dummy variable indicating whether the customer disconnects services in each month: 1=retain, 0=leave.	.84(.36)

Table 1B. Probability of Incoming Calls[a]

Sample Prob.		Not Call in	Call in with Tech Prob	Call in with Tran Prob
Starting Prob.		94.75%	4.97%	.28%
	Not Call-in	89.40%	9.45%	1.15%
Last period	Call-in with Tech Problem	71.91%	26.50%	1.59%
	Call-in with Tran Problem	80.19%	13.49%	6.32%

a. At t = 1, $\rho_1(D_{ikt-1})$ and $\rho_2(D_{ikt-1})$ are given by the starting probabilities .28% and 4.97%, respectively. From t = 2 forward, the conditional probabilities calculated from the sample are reported in the table.

Table 2. Comparative Advantages

	Onshore			Offshore		
	Overall	Transactional	Technical	Overall	Transactional	Technical
ALLOCATION	.84(.37)	.11[a](.32)	.89(.32)	.16(.37)	.03(.18)	.97(.18)
DUR	2.46(27.8)	6.39(17.61)	22.32(28.37)	37.62(24.69)	44.20(31.79)	36.28(24.45)
RET	.88(.31)	.91(.29)	.87(.31)	.83(.36)	.82(.38)	.84(.36)

[a]The percentage of questions handled by onshore centers that are transactional questions. SD is given in the parentheses.

Table 3. Model Comparison[a,b]

Models		Calibration Sample			Holdout Sample		
		Log L	BIC	AIC	Log L	BIC	AIC
Benchmark	1	-27175.4	54466.3	54408.8	-13455.2	27016.7	26968.3
Response	2	-23896.8	47949.0	47871.7	-11845.2	23833.5	23768.4
Models	3	-14473.3	29125.8	29036.5	-7182.3	14529.6	14454.5
Proposed	1 Segment	-16177.7	32462.9	32409.4	-8056.4	16211.8	16166.7
Response	2 Segments	-13061.7	26318.7	26221.4	-6422.7	13025.1	12943.3
Model	3 Segments	-23393.8	47082.5	46935.7	-11418.7	23108.9	22985.5

Notes: a. Benchmark model 1 is the proposed model without duration and waiting time, off-shore dummy, the interaction between off-shore and technical question dummy. Model 2 is the proposed model without off-shore dummy and the interaction term. Model 3 is the proposed model without the interaction term.
b. The Akaike and Bayesian information criteria (AIC and BIC) show that the four customer response models with two latent segments fit the data best (due to space constraints, we present only the results of the two-segment version for the benchmark models).

Table 4. Comparison with Sample Statistics

Statistics	Actual	Calibration Model
Percentage of allocations to offshore	.16	.18 (.03)
Average service duration in minutes	29.99	27.71 (3.67)
Average retention	.84	.86 (.03)

Table 5. Comparing Customer Demand for Service

Parameters		Coefficients (Std. Dev.)	
		Segment 1	**Segment 2**
Segment Membership		.663 (.005)*	.337 (.005)*[a]
	Intercept	1.065 (.018)*	.000 (.000) [b]
	EXP	.079 (.011)*	.000 (.000)
	RESIDENTIAL	-1.019 (.028)*	.000 (.000)
Log(Duration) (Regression)	Intercept	.508 (.031)*	5.225 (1.631)*
	Lagged log(Duration)	.101 (.007)*	.344 (.019)*
	TECHNICAL	.982 (.008)*	.139 (.001)*
	OFFSHORE	3.154 (.023)*	3.115 (.199)*
	OFFSHORE*TECHNICAL	-1.007 (.029)*	-1.401 (.113)*
	NCOMPUTER	.060 (.014)*	.410 (.094)*
	Variance	.526 (.010)*	.050 (.023)*
Retention (Binary Logit)	Intercept	7.488 (.304)*	.596 (.001)*
	SAT	6.581 (.378)*	1.910 (.030)*
	PRICE	-.049 (.005) *	-.466 (.001)*
	PROM	.142 (.025)*	-.013 (.794)
	COMP	-1.185 (.475)*	-.366 (.024)*
	Log(TENURE)	.889 (.498)*	1.273 (.132)*
	PENALTY	.468 (.136)*	.399 (.035)*
	OFFSHORE	-17.704 (7.553)*	-4.379 (1.223)*
	OFFSHORE*TECHNICAL	-3.135 (7.313)	7.114 (.111)*
	Log(Duration)	4.014 (1.193)*	4.733 (.110)*
	Log(Duration)^2	-2.264 (.248)*	-2.050 (.061)*
	WAIT	-1.486 (.416)*	-4.111 (1.312)*
	NCALLS	-1.333 (.081)*	-3.449 (3.514)
	FREQ_OFF	-46.702 (2.014)*	-1.440 (.124)*
	ACCUMDUR	-3.795 (.161)*	-1.462 (.200)*

[a] Following convention, we estimate all coefficients explaining segment membership together with the coefficients in the duration and retention equations and rely on model selection criteria to select the number of segments. We obtain the posterior membership using the logit formula based on the means of covariates (i.e. EXP, RESIDENTIAL) in the segment membership equation.

[b] For identification purposes, the coefficients for segment 2 are normalized to 0.

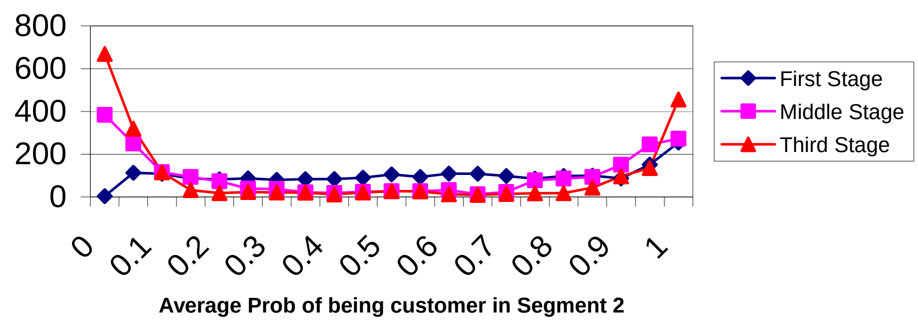
Table 6. Comparison of Actual and Proposed Allocation Strategies over the Entire Observation Period

	Actual			Proposed		
	Onshore	Offshore	Total	Onshore	Offshore	Total
Percentage of cases assigned	84%	16%	100%	81%	19%	100%
Percentage of technical questions	83%	17%	100%	80%	20%	100%
Percentage of customers in segment 2	66%	34%	100%	55%	45%	100%
Average service duration	20.46	37.62	29.99	14.22	18.67	16.32
Average waiting time	1.07	.75	.81	1.04	.76	.80
Percentage of customer retention at last period	88%	83%	84%	97%	89%	92%

Table 7. Performance Comparison across Optimization Frameworks

Optimization Framework	Accumulated Total Profits	Accumulated Total Costs	Retention Rate at Last Period
Observed Routine	108,984.0	717,204.5	84.00%
Alternative Nested Framework 1	153,214.0	668,619.0	85.26%
Alternative Nested Framework 2	163,879.6	661,692.9	88.42%
Alternative Nested Framework 3	164,346.8	680,782.3	89.07%
Alternative Nested Framework 4	166,797.2	679,539.3	91.26%
Proposed Framework	167,988.7	669,039.1	92.27%

Figure 1: Percentage of Customers in Segment 2 by Period



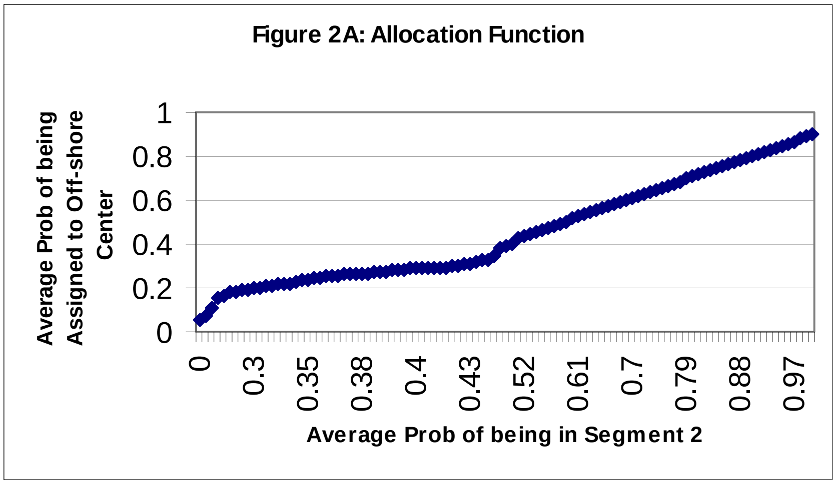


Figure 2A: Allocation Function
Average Prob of being Assigned to Off-shore Center
1
0.8
0.6
0.4
0.2
0
0
0.3
0.35
0.38
0.4
0.43
0.52
0.61
0.7
0.79
0.88
0.97
Average Prob of being in Segment 2

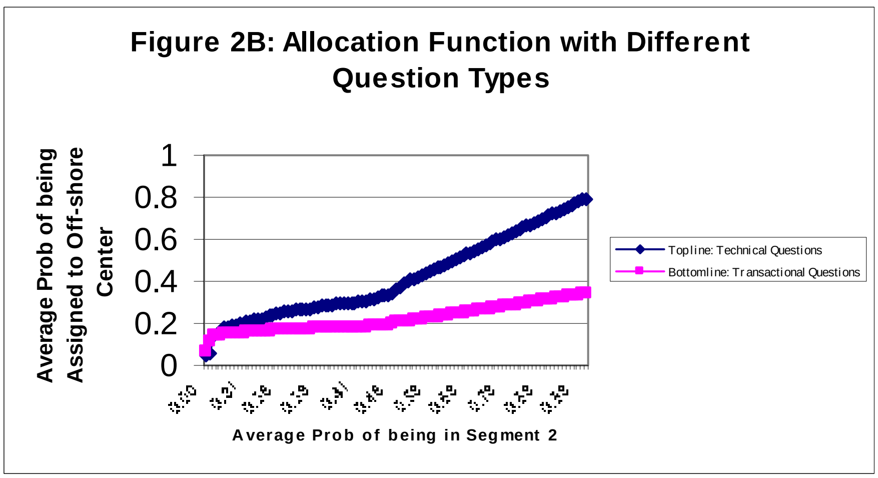


Figure 2B: Allocation Function with Different Question Types
Average Prob of being Assigned to Off-shore Center
1
0.8
0.6
0.4
0.2
0
Average Prob of being in Segment 2
Topline: Technical Questions
Bottomline: Transactional Questions

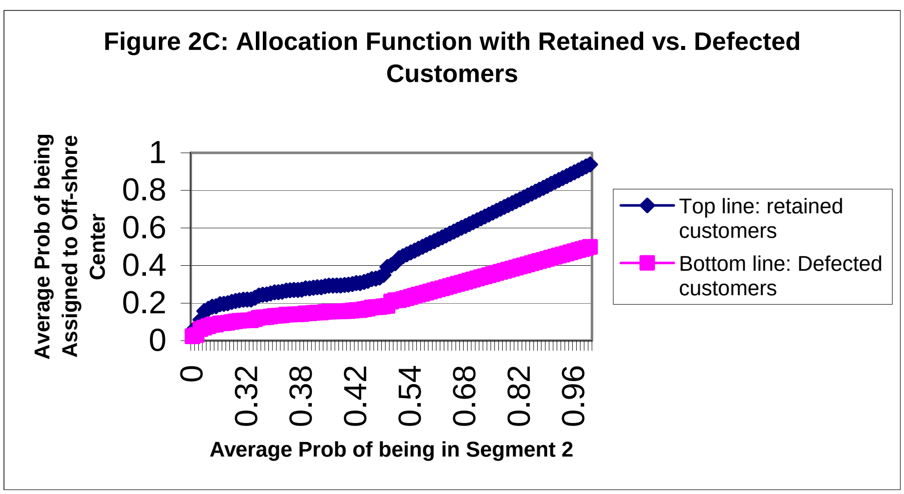


Figure 2C: Allocation Function with Retained vs. Defected Customers
Average Prob of being Assigned to Off-shore Center
1
0.8
0.6
0.4
0.2
0
0
0.32
0.38
0.42
0.54
0.68
0.82
0.96
Average Prob of being in Segment 2
Top line: retained customers
Bottom line: Defected customers

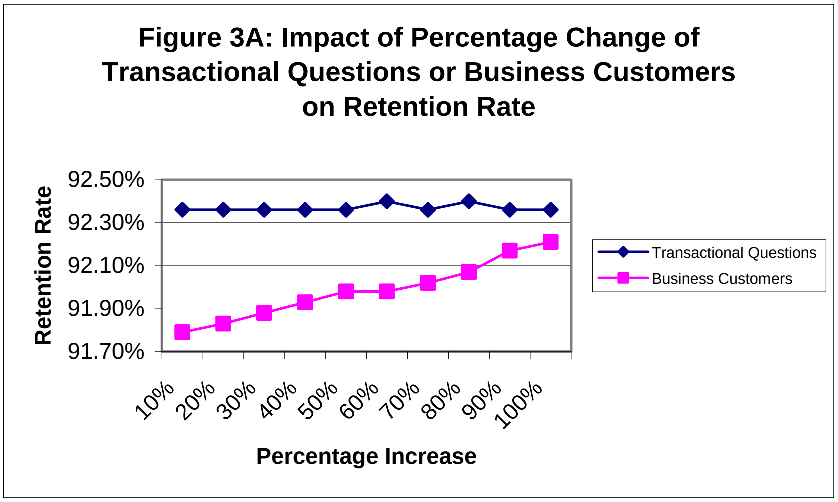


Figure 3A: Impact of Percentage Change of Transactional Questions or Business Customers on Retention Rate
Retention Rate
92.50%
92.30%
92.10%
91.90%
91.70%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Percentage Increase
Transactional Questions
Business Customers

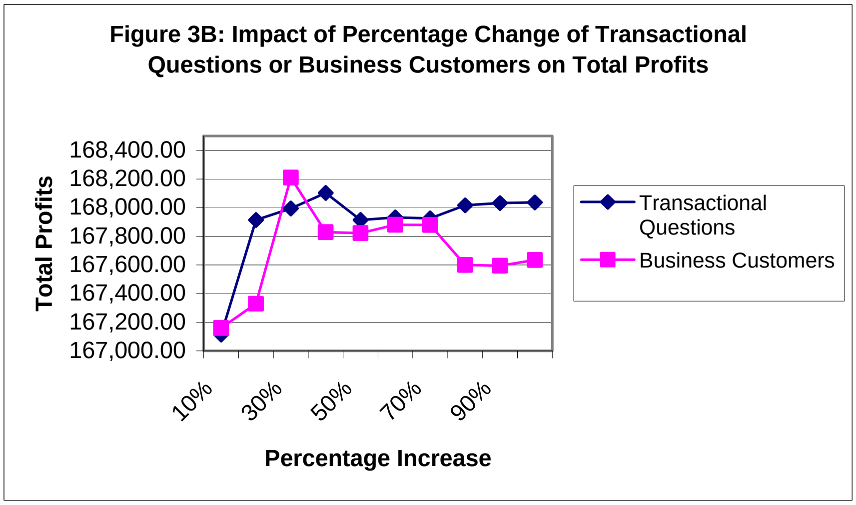


Figure 3B: Impact of Percentage Change of Transactional Questions or Business Customers on Total Profits
Total Profits
168,400.00
168,200.00
168,000.00
167,800.00
167,600.00
167,400.00
167,200.00
167,000.00
10%
30%
50%
70%
90%
Percentage Increase
Transactional Questions
Business Customers

Learning and Acting Upon Customer Information:

A Simulation-Based Demonstration on Service Allocations with Offshore Centers

Baohong Sun

Shibo Li

Web Appendix

Summary of Computational Algorithm in Simulations

We conduct simulations to derive optimal allocation decisions using our proposed dynamic programming framework in the following main steps.

Step 1: Read in the estimated parameters in the customer response model (Θ), the observed call history (D_{ikt}), and all the exogenous variables in the customer response models (Equations 1 and 4), and customer demographic variables (i.e., EXP and RESIDENTIAL).

Step 2: Start the customer loop. Within the customer loop, start the time loop for the customer.

Step 3: For each period of time, compute the segment membership probabilities $\Pr_{it}(m)$ (for the first period of time, we use $\Pr_{i0}(m)$ that is derived from the latent class estimates).

Step 4: Start the loop searching for waiting time for each call based on the approach described in Equation 16. In the simulation, we search for the best allocation rule and resulting waiting time through iteration. Starting from the waiting time ($WAIT_{jt}^{0}$) that we observe, we calculate a new waiting time $WAIT_{jt}^{1}$ using the derived allocation. Then $WAIT_{jt}^{1}$ serves as an input to determine retention, and we derive a new allocation rule. The iteration stops when $WAIT_{jt}^{s}$ converges such that waiting time is consistent with the proposed allocation rule.

Step 5: Simulate random draws for the error terms of customer's call duration and retention functions (Equations 1 and 4). Calculate expected service duration ($E[DUR_{ijkt}(m)| I_{it}, A_{it}]$) with

Equation 3. Calculate the probability ($\Pr_{it}(RET)(m)$) that consumer i of type m will stay with the firm at time t based on Equation 6. Calculate the firm's the average expected profits according to Equation 10 across all the random draws.

Step 6: Simulate the random draw for the error term of the firm's objective function and compute the average firm's utility across all draws (Equation 11).

Step 7: Derive the simulated optimal call center allocation decisions by solving the Bellman equation (Equation 15) using backward induction.

Step 8: Update the firm's perceived segment membership probability $\Pr_{it}(m)$ periodically according to Equations 7 and 8.

The simulations for other four optimization frameworks in Table 7 are similar for the proposed model.