M | MICHIGAN ROSS

Working Paper

The Value of Measuring Customer Satisfaction

Jihoon Cho
Stephen M. Ross School of Business
University of Michigan

Anohca Aribarg
Stephen M. Ross School of Business
University of Michigan

Puneet Manchanda
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper
Working Paper No. 1283
January 2017

The Value of Measuring Customer Satisfaction

Jihoon Cho, Anocha Aribarg, Puneet Manchanda*

July 2015
This version: January 2017

Abstract

A growing number of service firms now collect customer satisfaction ratings, along with objective service performance measures, for each service transaction. However, little is known about whether these two types of data are substitutes or complements, from both a conceptual and an applied point of view. This paper answers this question via the use of unique data consisting of individual-level cross-sectional and time-series measures of objective service performance, customer satisfaction, and purchase behavior. Using theory from the customer satisfaction literature, the data are applied to a two stage model of customer satisfaction and interpurchase time. The results suggest that the two sources of data provide complementary insights. In other words, customer satisfaction data provide information on business outcomes over and above that obtained from objective service performance data. The benefit of using and the cost of collecting these data are also quantified. The results are consistent across two different - quick service restaurant and auto rental - service industries, suggesting that they may be generalizable.

Keywords: *Customer Satisfaction, Service Quality, Expectation Disconfirmation, Performance Inconsistency, Selection Bias, Quick Service Restaurants, Auto Rental*

---

*Jihoon Cho is a doctoral student, Anocha Aribarg is Associate Professor of Marketing and Puneet Manchanda is Isadore and Leon Winkelman Professor and Professor of Marketing, all at the Stephen M. Ross School of Business at the University of Michigan. The authors are grateful to Xiaojing Dong, Fred Feinberg, Marc Fischer, Yesim Orhun, Adithya Pattabhiramaiah, S. Sriram, Ying Xie, (seminar) participants at the University of Michigan, the 2015 University of Cambridge Marketing Camp, Fudan University, University of Zurich, University of Technology Sydney, UT Dallas FORMS conference 2015 and INFORMS Marketing Science Conference 2013 for helpful comments. The authors are also grateful to two anonymous companies and the Wharton Customer Analytics Initiative for providing the data, as well as Melissa Hartz and Ben Adams for providing us with institutional details on the auto rental industry. This paper is based on the first essay of the first author's doctoral dissertation. All errors in the paper are our own. Please direct all correspondence to the first author at 701 Tappan St., Ann Arbor, MI 48109, USA or via email to *jihoonch@umich.edu*.

Firms in multiple industries use customer satisfaction as a measure of their service performance. Traditional customer-satisfaction surveys are expensive and time-consuming (requiring face-to-face, telephone or mail interviews) and therefore most firms carry out these surveys relatively infrequently (typically annually or even less frequently). However, with the emergence of Web 2.0 technology, customers can respond to a satisfaction survey with a click of a mouse, and their responses can be recorded and processed instantly. The ease of data collection has urged a growing number of service firms to solicit customers' responses to a satisfaction survey after each service encounter or transaction. Besides this, web-based transaction level satisfaction surveys also provide other benefits to firms - lower cost data collection and entry, easy access to all or most of their customers and the ability to present survey information in different formats (Couper 2000).

The large volume of such transaction-specific customer satisfaction surveys reported by companies that carry them out shows their prevalence. For example, Mindshare Technologies (a customer satisfaction survey specialist) reported in 2011 that it carried out 175,000 surveys every day, or more than 60 million surveys annually. ForeSee, an offshoot of the American Customer Satisfaction Index, carried out 15 million surveys in 2011 (Grimes 2012). Concomitant with the ability to carry out these surveys expeditiously, advances in information technology are increasingly providing service firms with the capability to monitor each service encounter on *objective* metrics easily and in a very cost effective manner. For example, UPS uses real-time delivery tracking (Lund and Marinova 2014), McDonald's monitors its drive-through service time (Hess, Ganesan, and Klein 2003) and airlines track the percentage of on-time flights (Grewal, Chandrashekaran, and Citrin 2010). These high quality, high frequency data have multiple benefits - they provide "objective" metrics, they can be generated without contact with customers and usually no third party (e.g., a customer satisfaction survey firms) needs to be involved in the data collection.

However, little is known about whether these two sources of data - customer satisfaction and objective service performance - are substitutes or complements, from both a conceptual and an applied point of view. In this paper, we leverage unique high-frequency data from two distinct service settings - quick service restaurants and auto rental - to answer this question. Unlike most previous customer satisfaction research that uses low-frequency, cross-sectional self-reported data, our data consist of individual-level cross-sectional and time-series measures of objective service performance, customer satisfaction and purchase behavior.

The use of these data has several benefits. First, observing objective service performance allows us to examine the additional value of customer satisfaction data relative to an objective benchmark. This is important as previous research has theorized that factors other than objective performance (e.g., brand trust, brand image, advertising, price image etc.) also impact customer satisfaction and repurchase behavior (Bolton and Lemon 1999; Rust, Lemon, and Zeithaml 2004). Second, our data avoid the problem of relationship inflation among variables induced by high common-methods variance (Fishbein and Ajzen 1975) that was prevalent in previous analyses where customer satisfaction ratings, objective service performance and purchase behavior were typically obtained from diverse channels (as opposed to from a single source). Lastly, the panel structure of our data allows us to control for a *within-individual* selection bias (as the ratio of completed surveys to transactions in such settings is relatively low) for such high-frequency satisfaction data. This bias can arise if the choice of whether to answer a satisfaction survey by a given customer is correlated with the quality of the service encounter. While limited previous research has looked at the across-individual selection bias (Godfrey, Seiders, and Voss 2011; Mithas, Krishnan, and Fornell 2005), there is virtually no work that has controlled for within-individual selection biases.

We answer our research question by modeling individual and transaction level outcomes - rating incidence, satisfaction rating and interpurchase time - using a two-stage model. In the first stage, we use a system of simultaneous equations with the first equation capturing drivers of customers' propensity to rate (a binary Probit model) and the second equation capturing the drivers of satisfaction rating (a linear regression model). In the second stage, we model customers' interpurchase time as a function of objective service performance and the aspects of customer satisfaction not explained by the objective service performance via a control function approach (i.e., via the use of the *residual*) (Petrin and Train 2010), along with other covariates, using a proportional hazard model. To allow for the heterogeneity in customer responses, we cast our model in a hierarchical Bayes framework.

Our results show that objective service performance affects both customer satisfaction and interpurchase time directly. More importantly, we find that the component of customer satisfaction that is not explained by the objective service performance has a significant impact on interpurchase time. This finding suggests that these two sources of data are complements and not substitutes. We obtain these results after correcting for within-individual selection and document the negative correlation between rating incidence and satisfaction rating. Our results also suggest that, in the context of transaction-level customer satisfaction data, within-individual selection likely to induce larger biases than across-individual selection. We also show that our results generally replicate via a set of robustness checks (e.g., using different measures of dependent and independent variables) and are generalizable across the two data sets from different service industries.

In summary, this paper makes the following contributions. First, at the conceptual level, it adds to the customer satisfaction literature by examining the interplay between objective service performance, customer satisfaction and actual purchase behavior. Specifically, it poses and answers the question as to whether objective service performance and customer satisfaction data are substitutes or complements. Second, it does so in the growing setting of individual-level times-series cross-sectional transaction-specific customer satisfaction and objective service performance data - a setting that has not received as much attention in the literature. This setting brings in unique challenges such as the possibility of bias in customer responses due to within-individual selection. Third, the results show that these two sets of data are complementary i.e., customer satisfaction data, measured frequently at the customer transaction level, do indeed provide additional value over and above the value obtained from objective service performance data. In addition, the within-individual selection bias is statistically and economically significant and needs to be controlled for. Fourth, from a managerial point of view, it shows that the complementary nature of customer satisfaction data can improve the prediction of business outcomes even when the firm has access to high-quality objective service performance data. The paper also illustrates, via a counterfactual analysis, how a firm can obtain a bound on the economic value of collecting customer satisfaction data. Finally, the replication of the results across two distinct and different service industries suggests that they could be generalized to other service settings.

The remainder of the paper is organized as follows. We first present an overview of the theoretical and methodological issues relating to service performance and customer satisfaction, based on the literature. Next, we present our model and estimation procedure. We then describe the institutional setting, the data and the operationalization of the variables across the two different industries. We next present our estimation results, a series of robustness checks, and results from policy simulations. Finally, we conclude with a discussion of our key findings, research limitations, and directions for future research.

CONCEPTUAL BACKGROUND

Customer Satisfaction: Its Measurement and Antecedents

Previous research has defined customer satisfaction as a post-purchase evaluation of a product or service performance resulting from the customer's comparison of the actual performance in relation to pre-purchase expectation (Churchill and Surprenant 1982). Customer satisfaction has been conceptualized either as an overall attitude towards a brand or product (e.g., see the description of the American Customer Satisfaction Index - ACSI - in Fornell et al. 1996) or a transaction-specific reaction (Oliver 1980; Keiningham et al. 2014a; Knox and Van Oest 2014). The former refers to the customer's cumulative attitude based on all encounters and experiences with the organization. The latter references customer satisfaction with a specific, discrete service encounter and has typically been measured by asking survey participants to consider the most recent experience they had (Olsen and Johnson 2003). In this research, we focus on the latter obtained via online surveys after each service encounter, a practice that has become prevalent among firms' satisfaction tracking programs. The popularity of transaction-specific measures stems from its ability to capture changes in customers' perception of service over time (Keiningham et al. 2014a).

The general consensus in the marketing literature is that satisfaction is a function of disconfirmation, the discrepancy between performance and expectation (Anderson and Sullivan 1993; Oliver 1980; Parasuraman, Zeithaml, and Berry 1988; Weaver and Brickman 1974).[1] Previous literature has found consistently that positive disconfirmation relative to expectation (i.e., performance exceeds expectation) increases customer satisfaction and negative disconfirmation decreases customer satisfaction (Oliver 1980). Some research also reports asymmetric disconfirmation effects where the negative disconfirmation effect is stronger than the positive counterpart (Anderson and Sullivan 1993; Gijsenberg, van Heerde, and Verhoef 2015; Knox and Van Oest 2014).

There exist different approaches to operationalizing the disconfirmation construct. First, early research measured the objective discrepancy between expectations and performance

[1] Previous research has found mixed or weak effects of expectation and perceived performance on customer satisfaction (e.g., Churchill and Surprenant 1982; Voss, Parasuraman, and Grewal 1998).

outcomes in an experimental setup to derive a difference score (Weaver and Brickman 1974). Second, Parasuraman, Zeithaml, and Berry (1988) propose a multi-item scale called SERVQUAL to measure perceived service performance where disconfirmation is derived as a difference score between perceived performance and performance expectation ratings on different service aspects. To maintain the independence between the expectation and disconfirmation constructs, other research captures individuals' summary judgment of overall disconfirmation using a rating scaled anchored at "better than expected" and "worse than expected" (Churchill and Surprenant 1982; Oliver 1980). It is also possible that these self-reported measures may not match individuals' pre-consumption expectations due to cognitive dissonance, assimilation, or contrast (Oliver 1977). Our operationalization of disconfirmation is in line with the objective discrepancy approach with the novel feature that it is based on observational data.

In addition to disconfirmation, later research has also emphasized the importance of consistent service performance in maintaining customer satisfaction. For example, McCollough, Berry, and Yadav (2000) conduct scenario-based experiments to show that customer satisfaction is lower after service failure and recovery (even with high-recovery performance) than in the case of consistent error-free service. Rust et al. (1999) demonstrate that it is not necessary to exceed customer expectations to increase preference and receiving an expected level of bad service does not reduce preference. The reason is that despite their desire for better-than-expected service, customers also prefer *consistent* service performance.[2] This paper adds to the literature on the antecedents of satisfaction by showing how disconfirmation and performance inconsistency derived from objective service performance affect transaction-specific satisfaction ratings over time.

[2]Under very specific conditions, increased performance inconsistency can lead to positive outcomes (Bolton, Lemon, and Bramlett 2006; Sriram, Chintagunta, and Manchanda 2015).

Impact of Satisfaction on Purchase Behavior

Satisfaction research has consistently shown the impact of customer satisfaction on purchase intention (Anderson and Sullivan 1993) and downstream business outcomes such as service usage (Bolton 1998; Bolton and Lemon 1999), customer retention (Mittal and Kamakura 2001; Seiders et al. 2005; Voss et al. 2010), share of customer wallet (Cooil et al. 2007; Keiningham et al. 2014a; Van Doorn and Verhoef 2008), and firms' financial performance (Fornell, Rust, and Dekimpe 2010; Grewal, Chandrashekaran, and Citrin 2010; Luo, Homburg, and Wieseke 2010). In contrast to linking customer satisfaction to observed customer behavior or firm performance, the use of perceived service performance, customer satisfaction and purchase intention based on self-report information gathered from the same survey is likely to inflate the relationship(s) among these constructs (i.e., high common-methods variance) (Fishbein and Ajzen 1975).

To establish the relationship between customer satisfaction and business outcomes, the vast majority of studies rely on measuring customers' attitude-based satisfaction from one or multiple cross-sectional surveys and aggregate business outcomes (Rust and Zahorik 1993; Van Doorn and Verhoef 2008; Verhoef 2003) (See Table 1 for the summary). For example, Van Doorn and Verhoef (2008) administered three consecutive annual surveys to investigate the impact of satisfaction on customer share of wallet. In contrast, our research matches satisfaction ratings to each service transaction and hence better track changes in firms' performance in response to customer satisfaction. Recent research links transaction-specific satisfaction to self-reported share of wallet for some randomly chosen transactions in a retail context (Keiningham et al. 2014a). However, this paper does not examine the role of objective service performance and disregards within-individual selection. Keiningham et al. (2014b) also study how service severity (i.e., negative objective performance) affects satisfaction and subsequently market share but using aggregate cross-sectional data. Finally, our work is also related to other research that focuses on securing different objective service performance measures to study their relationships with actual business outcomes (e.g., Bolton, Lemon, and Bramlett 2006; Knox and Van Oest 2014; Lund and Marinova 2014; Sriram, Chintagunta, and Manchanda 2015). However, this research does not study transaction-specific customer satisfaction in the same framework.

[Table 1 about here.]

Selection Bias in Satisfaction Ratings

In general, not all customers reply to satisfaction surveys, opening up the possibility of the results being biased by self-selection. Failing to account for this selection bias that arises from systematic survey nonresponse will lead to biased inferences with regard to the observable factors that drive the outcome variable of interest. To address this concern, some previous customer satisfaction literature has controlled for the effect of unmeasured characteristics related to the selection process. Godfrey, Seiders, and Voss (2011) take into account *across-individual* selection by first modeling customer's propensity to be included in the satisfaction survey and then using the obtained inverse Mills ratio as a control variable that links customer satisfaction to repurchase. Mithas, Krishnan, and Fornell (2005) use a propensity score matching approach to control for the selection bias.

In transaction-specific survey settings, each individual can be solicited for her ratings multiple times in a year. This frequent surveying could lead to the decision to participate becoming a function of the service quality. For example, a consumer may rate a provider only on occasions when she had an unsatisfactory service experience. Thus, the decision to rate and the rating itself will not be independent due to the presence of a *within-individual* selection bias. The results from any analysis that does not explicitly correct for this bias could be misleading for the firm. In this paper, we test for the existence and magnitude of this bias via our model. To the best of our knowledge, our research is the first to address the issue of within-individual selection process in the satisfaction literature.

MODEL AND ESTIMATION

Model Specification

We model satisfaction rating incidence, satisfaction rating and interpurchase time as three separate but related processes, by constructing a set of simultaneous equations at the individual consumer level. In the first stage, we model two individual-level decisions at each service encounter: whether to provide satisfaction ratings and if so, what rating to give. We use a binary probit and a linear regression to model these two decisions respectively, while allowing for the errors across these two equations to be correlated. This allows us to explicitly control for a *within-individual* selection bias arising from factors that affect both decisions but are unobserved by researchers (see Narayanan and Manchanda (2012) for a similar situation in a different institutional setting). For customer i's receiving service from location j on service occasion t, the system of equations is specified as follows:

$$(1) \quad INC^{*}_{ijt} = \alpha_{0i} + \alpha_{\mathbf{1i}} X^{INC}_{ijt} + \epsilon^{INC}_{ijt},\ INC_{ijt} = 1 \text{ where } INC^{*}_{ijt} > 0,\ INC_{ijt} = 0 \text{ otherwise}$$

$$(2) \quad SAT_{ijt} = \beta_{0i} + \beta_{\mathbf{1i}} X^{SAT}_{ijt} + \beta_{2i} IMR_{ijt} + \epsilon^{SAT}_{ijt}$$

where INC^{*}_{ijt}, INC_{ijt} and SAT_{ijt} represent the underlying latent variables representing customer i's decisions of whether to rate, rating incidence, and overall satisfaction score, respectively. IMR_{ijt} illustrates the inverse Mills ratio generated from Equation 1. $X^{INC}_{ijt} = \{DIS_{ijt}, VAR_{ijt}, CPN_{ijt}, NTR_{ijt}\}$ and $X^{SAT}_{ijt} = \{DIS_{ijt}, VAR_{ijt}, CPN_{ijt}\}$ are sets of explanatory variables for each equation. Note that the residuals from Equation 2, $r^{SAT}_{ijt} = SAT_{ijt} - (\beta_{\mathbf{1i}} X^{SAT}_{ijt} + \beta_{2i} IMR_{ijt})$, are retained and used as a proxy for customer satisfaction based on factors other than objective service performance.

In the second step, we model the probability of purchase conditional on interpurchase time with a semi-parametric survival model (Cox 1975), incorporating objective service performance measures and customer satisfaction. Specifically, our approach focuses on a daily purchase decision, that is, customers decide every day whether they plan to purchase as a function of the timing of their last purchase and transaction details at previous service encounter. This model specification treats the no-purchase days for each customer as the survival days, whereas it treats the purchase days as the failure days (for a comparison of alternative specification, see Manchanda et al. 2006). This semi-parametric approach is appealing as it does not require the specification of the underlying purchase time distribution (Gupta 1991). Let $h(\tau|X_{\tau}^{INT})$ denote the hazard rate at time τ for an individual having covariate values X_{τ}^{INT} at time τ. This hazard rate is assumed to take the form:

$$(3) \quad h(\tau|X_{ij\tau}^{INT}) = h_0(\tau)\exp(\gamma_{1i}r_{ij\tau}^{SAT} + \gamma_{\mathbf{2i}}X_{ij\tau}^{INT}),$$

In the above expression, $h_0(\tau)$ represents a constant baseline hazard that corresponds to interpurchase time between purchase occasion t and $t-1$, instead of calendar-time, τ. $X_{ij\tau}^{INT} = \{DIS_{ij\tau}, VAR_{ij\tau}, CPN_{ij\tau}, INT_{ij\tau}, AMT_{ij\tau}\}$ indicates a set of covariates that enter in the proportional hazard formulation multiplicatively. As such, $\gamma_{\mathbf{i}} = \{\gamma_{1i}, \gamma_{\mathbf{2i}}\}$ can be viewed as the individual-specific proportional effect of customer satisfaction and X_{τ}^{INT} on the hazard rate. The exponential function simplifies the estimation of $\gamma_{\mathbf{i}}$ as no constraints need to be imposed to ensure nonnegativity (Helsen and Schmittlein 1993).

Following the previous literature (Weaver and Brickman 1974) that takes a more objective approach, we operationalize disconfirmation, DIS_{ijt}, as the difference between current objective service performance and prior customer expectations of the performance. Given the availability of multi-period panel data, we specify the evolution of customer expectation to follow an anchoring and adjustment process (Nerlove 1958) and derive it as a function of objective service performance that varies over time.[3] Previous literature reveals that this adaptive expectation framework provides a reasonable fit to observed customer behavior (e.g., Erdem 1998). In this setup, the greater the weight on the objective performance, the more significant the effect of immediate past experience on current expectation, or the more adaptive the expectations (Johnson, Anderson, and Fornell 1995).

$$(4) \quad DIS_{ijt} = PERF_{ijt} - EXP_{ijt} \text{ where } EXP_{ijt} = \delta\, PERF_{ijt-1} + (1-\delta)\, EXP_{ijt-1}$$

In the above expression, $PERF_{ijt}$ and EXP_{ijt} represent objective service performance and customer expectation of the performance, respectively. The parameter δ is an empirically derived factor that determines the relative weights assigned to the prior expectation and the current service performance. To determine the exponential smoothing factor, δ, we perform a grid search (e.g., Fader, Lattin, and Little 1992). We let δ vary from 0 to 1 with increments of 0.1. Note that previous research has typically constructed the disconfirmation variable using survey-based measures (as opposed to objective measures as we use) of expectation (Boulding et al. 1993; Parasuraman et al. 1988).[4]

Service performance inconsistency, VAR_{ijt}, is operationalized as an individual-level cumulative standard deviation of delivery time up to the current service encounter (e.g., Sriram, Chintagunta, and Manchanda 2015). In addition, we include customer coupon redemption,

[3]Note that our approach to derive customer expectations helps us circumvent the mere-measurement effect. Previous research argues that prompting customer expectations sensitizes negative feelings. For example, Ofir and Simonson (2007) show that customers who had been solicited their expectations by the researchers gave the store lower post-shopping satisfaction ratings than did those who had not.

[4]For example, Boulding et al. (1993) proposes two different classes of survey-based expectation measures - "will" expectation and "should" expectation. Will expectation is specified as a weighted average of prior expectations and actual service performance (closer to our measure), while should expectation is updated only when the firm's service performance exceeds a customer's prior should expectations. The authors find that while will expectation increases perceived quality, should expectation decreases it. Given the use of survey-based expectation and their focus on the relationship between expectation and perceived quality, we cannot compare our results directly with theirs. Nonetheless, we created another measure of expectation instead of our current one (i.e., will expectation) where we updated expectation only when delivery time was shorter than expectation (i.e., should expectation). We find disconfirmation based on should confirmation to also decrease customer satisfaction.

CPN_{ijt}, and log-transformed dollar purchase amount, AMT_{ijt}, as control variables. The error terms in Equation 1 and 2, ϵ_{ijt}^{INC} and ϵ_{ijt}^{SAT}, are assumed to have a multivariate normal distribution with mean vector of zero and covariance matrix of $(1, 1; \rho_{12})$. This error structure explains the correlation between unobserved components in customer rating behavior and controls for the *within-individual* selection problem. We fix the scale of the latent utilities by imposing the restriction that the variances of ϵ_{ijt}^{INC} and ϵ_{ijt}^{SAT} be unity.

To address the potentially endogenous relationship between rating incidence and satisfaction rating, we use the number of transactions since the last time customer i provided a satisfaction rating, NTR_{ijt}, as an exclusion restriction. This variable works as an excluded variable as it has been shown to impact participation in the survey but not the rating itself e.g., frequent solicitations to answer (similar) surveys has been shown to lower participation (Bickart and Schmittlein 1999). This is because with increasing contact, respondents' overall attitudes toward the survey become less favorable, and they feel that the opportunity to provide their opinions in a survey is not a "rarity" and, therefore, not a valuable experience (Groves, Cialdini, and Couper 1992).[5]

We use random coefficients to control for unobserved heterogeneity at the individual level. Specifically, we cast our model in a hierarchical Bayesian framework to obtain individual-specific parameters in the rating incidence, satisfaction rating, and interpurchase time equations. Finally, it is also possible that there may be unobserved factors related to store characteristics (e.g., store size, the date when the store opened etc.) that systematically affect the dependent variables of interest. However, we expect that such differences in store characteristics will be captured by the individual-specific random intercepts as the orders from each customer are almost always confined to a certain store based on his/her address.

[5]Note that for this variable to act as the excluded variable, the amount of serial correlation for rating incidence and satisfaction rating should not be high simultaneously. This condition holds as the serial correlation for rating incidence is 0.09 for the entire sample while that for satisfaction rating is 0.36.

Estimation

In order to estimate our proposed model, we first fit the satisfaction rating model (Equation 2) together with the rating incidence model (Equation 1), employing the Heckman selection framework (Heckman 1979). Specifically, we obtain the inverse Mills ratio, $IMR_{ijt} = \phi(\alpha_{0i} + \alpha_{\mathbf{1i}} X_{ijt}^{INC}) / \Phi(\alpha_{0i} + \alpha_{\mathbf{1i}} X_{ijt}^{INC})$ from (Equation 1) and use the value as a control variable in Equation 2 to address within-individual selection. With the parameter estimates from Equation 2 in hand, we obtain the residuals, $r_{ijt}^{SAT} = SAT_{ijt} - (\beta_{0i} + \beta_{\mathbf{1i}} X_{ijt}^{SAT} + \beta_{2i} IMR_{ijt})$, that capture the information contained in the satisfaction rating over and above objective service performance and coupon redemption. Using the set of purchase occasion observations C, we estimate a proportional hazard model (Equation 3) with a log-likelihood function (Cox 1975; Golder and Tellis 2004) as follows:

$$(5) \qquad LL_i(\gamma_i) = \sum_C \left[\gamma_{1i} r_{ij\tau} + \gamma_{\mathbf{2i}} X_{ij\tau} - \log \sum_{i' : INT_{i'j\tau} \geq INT_{ij\tau}} \exp(\gamma_{1i'} r_{i'j\tau} + \gamma_{\mathbf{2i'}} X_{i'j\tau}) \right]$$

We capture the heterogeneity in $\{\alpha_{\mathbf{i}}, \beta_{\mathbf{i}}, \gamma_{\mathbf{i}}\}$ across individuals by allowing them to be distributed multivariate normal with mean $\{m_\alpha, m_\beta, m_\gamma\}$ and variance $\{V_\alpha, V_\beta, V_\gamma\}$. The hyperparameters $\{m_\alpha, m_\beta, m_\gamma\}$ and $\{V_\alpha, V_\beta, V_\gamma\}$ are distributed multivariate normal and inverse Wishart, respectively. We draw these parameters by constructing an MCMC chain where a Gibbs sampler (for parameters whose full conditionals can be derived) is used in combination with data augmentation (e.g., Kai 1998) for the purchase incidence equations and a Metropolis-Hastings sampler for the interpurchase time equation. As our MCMC sampler is fairly standard, we omit the details for in the interest of brevity (though these are available from the authors on request).

STUDY 1: QUICK SERVICE RESTAURANT INDUSTRY

Institutional Background

We obtained data from a large American company in the quick service restaurant industry. The company has an international presence and operates own stores as well as franchises. The food delivery context is of interest to us because the service component, especially delivery time, of the transaction influences both customer satisfaction and purchase behavior (Verma, Thompson, and Louviere 1999). Timely service has been widely accepted as a key to success in the service industry because it is the first interaction in the sequence of experiences that customers have with the firm (Bitner 1992). Firms can also improve service time as a means of differentiation based on convenience (Lund and Marinova 2014). The nature and order of these experiences thus can have an impact on overall service satisfaction (Chase and Dasu 2001). This is true in our setting as well with the company's managers confirming that delivery time is the main determinant of service performance for their customers. As a result, the company has made a significant investment in tracking food preparation and delivery time. Specifically, the company requires each store to record four different time stamps for each order: when the order is placed (TS1), when the order comes out of the oven (TS2), when the driver leaves the store (TS3), and when the driver returns to the store (TS4). The delivery time for each order is calculated to be $[(\text{TS4} - \text{TS3})/2 - \text{TS1}] + 2$ minutes.[6] Based on both the previous literature and our specific setting, we use delivery time as the key objective service performance measure in our analysis. Other measures of service performance (e.g., number of service failures, telephone CSR service quality, frontline employee interactions, product quality) are also potential determinants of customer satisfaction and business outcomes. However, in this industry (and in our setting), none of these are obtained at each transaction level.

In addition to investing in its own tracking, the company has also invested in making the service experience transparent to the customer. Specifically, the company provides its

[6]The two-minute addition is based on a calibration exercise carried out by the company. We are able to replicate the results if we subtract two minutes from each delivery time.

customers with a unique online order experience through its online "order tracker." After an order (online, phone, or walk-in) is placed, the customer can monitor the status of the order directly from the company's website - she can track when the food preparation is complete (at the store) and when the order gets sent out for delivery. On the website the customer is prompted to fill out a five-point scale satisfaction survey with respect to her order. As customers make these satisfaction assessments immediately after the delivery, we assume that judgments of the service encounter are affected by only the actual service performance experienced in that transaction (e.g., Zhang and Kalra 2014). The survey consists of six questions as below:

- Q1: How likely are you to recommend us to your family and friends?
- Q2: How fast and nice was your phone order?
- Q3: How would you rate your online ordering experience?
- Q4: How would you rate your delivery experience with driver?
- Q5: How would you rate your carryout experience?
- Q6: How would you rate the quality of your order?

Data Description

The data span a total of 743,609 delivery orders from 99,156 unique customers (households)[7] who provided satisfaction ratings at least once during the sample period at 625 stores in Texas and Virginia from January to December 2011.[8] The transaction details include store ID, order date, order ID, delivery time, customer ID, coupon redemption, pick-up method (carryout vs. delivery), purchase amount and satisfaction ratings. Given our interest in

[7]As our data are at the household level, we cannot separately identify whether repeat purchases by the household represent true repeats by the same person or are new purchases by someone else in the household. We therefore use the term "customer" and "household" interchangeably.

[8]The overall proportion of delivery orders is 57.8%. We also have an additional 6,655,320 delivery orders from 1,136,700 customers who did not provide any ratings during the sample period. We use the data from these "non-raters" to check whether there is a potential *across-individual* selection bias in the "Robustness Checks" section.

delivery time as the objective performance measure, we restrict our attention to delivery orders. We do not observe substantial within-household heterogeneity in ordering methods. Approximately 90% of customers in the data use the same method of order during the sample period (27.5% of carryout-only and 62.1% delivery only). In addition, within-household heterogeneity in store choices is also minimal. This is because once a customer enters his/her address online, the website automatically locates the stores that are closest to his/her address. This results in only 4.6% of all the transactions where customers order from different stores over time. As we derive performance inconsistency from the observed objective performance (i.e., delivery time), we also need to observe at least three observations per customer, e.g., we need two observations in T_0 and T_1 to compute performance inconsistency and link it to purchase behavior in T_2. We thus limit the sample to 484,440 transactions from 74,080 customers who purchased three or more times. As can be seen from Table 2, there is no significant difference in transaction details and behavior between the households in sample with at least three purchases and the entire sample.

[Table 2 about here.]

From the six questions in the online survey, we use the mean response across Q3 (How would you rate your online ordering experience?) and Q4 (How would you rate your delivery experience with driver?) on a 5-point scale to construct a measure of overall transaction-specific customer satisfaction. Morgan and Rego (2006) show that this measure of overall transaction-specific satisfaction provides greater value in predicting future business performance and is conceptually distinct from the Net Promoter Score or average number of recommendations. The mean and median of this measure over the sample period are 4.68 and 5, respectively. We do not use Q1 and Q6 because the former is more like a recommendation measure (like the Net Promoter Score) and the latter is about overall perceived quality. Q2 and Q5 appear to capture responses about phone and carryout order, which are not relevant to our study context.[9] Figure 1 shows the distribution of our overall satisfaction measure and its correlation with delivery time. As shown, customers' evaluations are skewed towards the highest score and our focal objective service performance - delivery time - is negatively correlated with this overall satisfaction measure (the correlation coefficient is -0.19 and a regression of overall satisfaction on delivery time shows that the latter has a significant and negative effect).

[Figure 1 about here.]

Customer satisfaction ratings were provided in 2.2% of all transactions. While this may seem low, this is consistent with industry statistics (based on feedback from the data provider). Conditional on individuals providing a rating during the sample period (i.e., "raters"), participation went up quite dramatically with 21.6% of transactions being rated, i.e., 1.8 times per individual on average. Table 3 presents the summary statistics (1) from transactions with satisfaction ratings and (2) from transactions without satisfaction ratings. Overall, there are no statistically significant differences in transaction details across the two different samples. However, interpurchase time is 55.6% longer and the number of transactions since the last rating is 25.4% smaller for transactions with satisfaction ratings relative to those without ratings. In order to exploit the panel nature of the data and to correct for within-individual selection, we use the transactions both with and without satisfaction ratings for these households in our analysis.

[Table 3 about here.]

As noted earlier, our chosen sample excludes "non-raters" (i.e., households that had not participated in the satisfaction survey even once during our data period). Table 4 presents the key variables of the transactions made by (1) "raters" and (2) "non-raters."

[9] Over 90% of Q2 and Q5 are missing mostly because phone-order/carryout customers do not seem to go to the firm's website to track their order status even if they have access to the tracker.

The descriptive statistics suggest that the differences between "raters" and "non-raters" on the key metrics are not as substantial as those between transactions with and without ratings from customers who rated at least once. As a robustness check, however, we later try to correct for across-individual selection to detect its presence and compare its magnitude to that of the within-individual selection (See the "Robustness Checks" section).

[Table 4 about here.]

Results

Customer Rating Behavior. In this section, we show the results from the proposed model of customer rating behavior that includes both the decision to rate and the actual rating, conditional on the rating decision. Table 5 reports the results from two different specifications: (1) the Heckman selection framework of rating incidence and satisfaction rating to control for within-individual selection and (2) a null model where we ignore within-individual selection (i.e., only use a linear regression model for satisfaction rating). In both specifications, we account for unobserved customer heterogeneity using the hierarchical Bayes framework. Model 1 is the proposed model and Model 2 is the null model.

[Table 5 about here.]

Overall, we find that objective service performance does have a significant impact on customer satisfaction rating. The parameter estimates in Table 5 suggest that both disconfirmation and performance inconsistency are key determinants of customer satisfaction rating. First, with greater disconfirmation (delivery time is longer than expected) customers are more likely to participate in surveys and provide lower satisfaction ratings. These results confirm findings from the previous literature that proposes a significant relationship between disconfirmation and customer satisfaction (e.g., Oliver 1980). Furthermore, inconsistent service performance significantly decreases both survey participation and customer satisfaction, suggesting that minimizing customer uncertainty plays an important role in maintaining customer satisfaction.

Model 1 confirms that there is a selection bias in within-household ratings over time as the coefficient of the inverse Mills ratio is significant and negative, and as such, it is important to correct for the within-individual selection. The negative coefficient suggests that customers are less likely to provide a rating when they feel satisfied with the service they received.[10] Note that ignoring this correlation results in 1.8% of underestimation of disconfirmation elasticity and 4.5% overestimation of performance inconsistency elasticity to customer satisfaction. The number of transactions since the previous rating - the proposed instrument to identify the selection process - significantly increases customers' participation in satisfaction survey. The more recently customers rated, the less likely they are to provide ratings again.

Customer Purchase Behavior. Next we focus on whether customer satisfaction provides additional information over and above the information present in objective service performance. Specifically, we link the residuals from Equation 2, along with disconfirmation and performance inconsistency, to the probability of purchase conditional on interpurchase time, using only those observations for which we have satisfaction ratings. We then compare the proposed model with a series of alternative models in order to answer our research question. Table 6 reports the results.

[Table 6 about here.]

The key result from the proposed model is that the residual has a positive impact on the purchase probability, after controlling for objective service performance.[11] This implies that

[10]Note that the negative coefficient of the inverse Mills ratio can also potentially be driven by dissatisfied customers participating in the survey. However, the data suggest that most of the participation effect associated with satisfaction rating is attributable to satisfied customers. For example, the negative effect of delivery time on customer satisfaction is not significant when customers provide relatively lower ratings, while delivery time significantly increases survey participation.

[11]Note that this direct impact of customer satisfaction on purchase might be attributed to the mere-

customer satisfaction and objective service performance are complements, not substitutes. In addition, inconsistent service performance decreases the probability of purchase. The availability of a coupon also decreases the probability of purchase. Finally, lagged interpurchase time tends to lower the probability of purchase.

The proposed model uses constructed and/or transformed measures such as the residuals from the satisfaction equation, disconfirmation, performance inconsistency etc. However, firms could work directly with the raw subjective and objective measures. Models 2 and 3 use these measures directly. The results from Model 2 suggest that conditional on customer's decision to rate, the satisfaction rating (the raw subjective measure of service performance) positively impacts purchase behavior - satisfied customers are more likely to purchase, compared to the dissatisfied customers. The results from Model 3 present a strong direct relationship between the probability of purchase and delivery time (the raw objective measure of service performance). Model 4 uses only disconfirmation and performance inconsistency (without the residuals) measures and the results are qualitatively similar to those from Model 1. Model 5 ignores the theoretically motivated constructs of disconfirmation and performance inconsistency, using the delivery time along with the residual. Turning to model fit, we find that our proposed model provides the best predictive performance with respect to root mean square error (RMSE) and mean absolute error (MAE).

Collectively, the results highlight the value of collecting both objective and subjective measures of performance. The combined set of measures helps link objective service performance to purchase behavior and identify the effect of customer satisfaction over and above that of objective service performance. These findings rationalize this firm's decision to continue collecting customer satisfaction data even when they have access to high-quality

measurement effect where measurement of customer intentions or customer participation in surveys could positively influence customer repeat purchase behavior (Dong, Janakiraman, and Xie 2014). However, this explanation is unlikely in our context because the correlation between survey participation and interpurchase time (0.16) is in the opposite direction as suggested by mere-measurement literature.

objective service performance data, as the two sets of data are complementary. In Study 2, we check to see if we replicate this finding. Finally, in the section titled, "The Effect of Changes in Service Performance," we show how firms can compute the economic value of collecting customer satisfaction data and provide an estimate for the firm in question.

Robustness Checks

In this section, we report results from a series of robustness checks. First, we investigate the impact of objective service performance and customer satisfaction on purchase amount instead of the probability of purchase conditional on interpurchase time. Second, we test alternative measures of customer expectation to calculate disconfirmation. Third, we explore the asymmetric effect of disconfirmation on customer satisfaction. Finally, we examine the relative importance of across-individual and within-individual selection biases.[12]

Customer Purchase Amount. We first examine whether the results are robust to an alternative business outcome - the dollar amount of each order. Similar to the approach in Equation 3, we treat the residuals from the customer satisfaction equation (Equation 2) as an independent variable in a linear regression with the dependent variable being the log-transformed dollar purchase amount, which has been used extensively in marketing for modeling sales and expenditure (Blattberg and Neslin 1990). The results are reported in Table 7 (Column (1)). Similar to the findings in Table 6, both objective service performance and the residuals significantly increase the dollar purchase amount of each order. These results confirm the role of satisfaction ratings in providing incremental value in predicting another aspect of consumer purchase behavior in addition to interpurchase time.

[Table 7 about here.]

[12] We are also able to run a robustness check where we can explicitly control for *product quality* for a very small sample of transactions. We replicate many of our main results, including those pertaining to the impact of disconfirmation and performance inconsistency. Details are available from the authors on request.

Alternative Measures of Customer Expectation. Throughout the paper we operationalize customer expectation as an exponentially smoothed average of service performance up to the previous service encounter. As a robustness check, we use two alternative measures of customer expectation: most recent service encounter and a simple moving average across all previous service encounters. The former is motivated by the fact that customers may find it hard to recall all their prior experiences perfectly and to construct a summary measure due to factors such as the high cognitive effort required for adjusting prior expectations, low involvement and low purchase frequencies (e.g., Mitra and Golder 2006). The latter is based on the assumption that all past service encounters contribute equally to customer expectation. This is in contrast to the exponential smoothing approach used in the proposed model which gives higher weights to service performance that occurs more recently. Based on these alternative expectation measures, we calculate disconfirmation as in Equation 4. As shown in Table 7 (Column (2)), the results using the most recent service encounter as customer expectation are generally consistent with our main results as higher disconfirmation (i.e., worse-than-expected service performance) increases survey participation, decreases customer satisfaction but does not directly affect probability of purchase. The results using the simple moving average are qualitatively similar (they are omitted for the sake of brevity).

Asymmetric Impact of Disconfirmation on Customer Satisfaction. Prospect theory (Kahneman and Tversky 1979) suggests that people are less influenced by absolute values of certain factors than by changes in these values. Moreover, such changes will have a stronger effect when they are negative as compared to positive. In light of this theory, previous service literature (e.g., Gijsenberg, van Heerde, and Verhoef 2015; Knox and Van Oest 2014) has documented that satisfaction is likely to be more sensitive to negative disconfirmation than positive disconfirmation. To test this asymmetric effect of disconfirmation, we follow previous satisfaction literature (Anderson and Sullivan 1993) and decompose the disconfirmation variable, DIS_{ijt}, into positive disconfirmation, PD_{ijt}, and negative disconfirmation, ND_{ijt} and re-run the proposed model with these two disconfirmation constructs:

(6) $ND_{ijt} = DIS_{ijt}$ and $PD_{ijt} = 0$, if $TIME_{ijt} >= EXP_{ijt}$

(7) $PD_{ijt} = DIS_{ijt}$ and $ND_{ijt} = 0$, if $TIME_{ijt} < EXP_{ijt}$

As shown in Table 7 (Column (3)), we find that customer satisfaction is significantly influenced by both negative and positive disconfirmation. Notably, consistent with previous literature, negative disconfirmation (i.e., worse-than-expected service performance) has a stronger impact on customer satisfaction than positive disconfirmation (i.e., better-than-expected service performance). In particular, the parameter estimate of negative disconfirmation is 14.7 times larger than that of positive disconfirmation (this difference is significant at p=0.01 level).

Selection Issue Across Individuals. The proposed model accounts for non-rated (for satisfaction) transactions via the selection equation within individuals. There is also the possibility that customers who have never rated are different from those who rated at least once, leading to an across-individual selection problem. Thus, if the firm acts on the satisfaction ratings, they may not be acting optimally with respect to their entire customer base. We estimate an *across-individual* selection model where we include both "raters" and "non-raters" in the sample. To match the same number of observations we use in the models reported in Table 5 and 6, we draw a sample of 74,080 customers from the population of all customers who made at least three purchases during the data period. As a result, some customers in the sample are "raters" (i.e., rated satisfaction at least once) and the others are "non-raters" (i.e., never rate in this period). We use the last transaction of each customer to create the sample of observations. This sample is very similar to cross-sectional survey data commonly used in the industry and the previous literature to obtain satisfaction ratings. Table 8 (Column (2)) reports the results of the model estimated on this sample. The results show that across-individual selection is not significant, while disconfirmation significantly affects survey participation and satisfaction ratings. These results suggest that the behavior of "raters" and "non-raters" are not significantly different and within-individual selection is a more crucial problem for firms to consider when analyzing transaction-specific satisfaction ratings in panel settings.

[Table 8 about here.]

The Effect of Changes in Service Performance

Our results show that customer satisfaction data contain useful information even after the inclusion of objective service performance data from a statistical point of view. In this section, we turn to quantifying the value of collecting customer satisfaction from a managerial point of view. To do so, we first set up a simulation to assess the impact of improved or delayed delivery time on customer satisfaction and interpurchase time. The delivery time is manipulated as follows: starting from the initial delivery time for each customer, we increased/decreased delivery time by $k\%$ each period (bounded by a minimum of 2 minutes and maximum of 2 hours, based on the data) where k ranges from 1 to 50. Next, based on the manipulated delivery time, we calculated the proportional changes in predicted interpurchase time, using the parameters estimated from the proposed model. These proportional changes represent the impact of service delay/improvement over time on interpurchase time. Note that these changes also imply a change in performance inconsistency - as the delivery time varies (up or down), performance inconsistency goes up and the firm will get penalized.[13]

[13] As both disconfirmation and performance inconsistency have an impact on customer behavior, our manipulation has to ensure a change in both. This is why we decrease or increase delivery for each successive purchase occasion in the manner described above. In contrast, a flat decrease or increase will not change performance inconsistency.

Thus, the impact of changed delivery times is a tradeoff between the direction of the change and the amount of additional performance inconsistency.

We examine the results from the above simulation in three different ways. First, we look at the impact on customer satisfaction as the objective service performance changes. Relative to a baseline case of no performance inconsistency, as can be seen from Figure 2a, a decrease in objective service performance (increased delivery times) lowers satisfaction significantly (up to 8.1%). On the other hand an increase in service performance (decreased delivery times) improves customer satisfaction, though to a lesser extent. In the case of a decrease in service performance, both disconfirmation and performance inconsistency lower satisfaction. In contrast, when service performance improves, disconfirmation improves customer satisfaction but that improvement is offset by the decrease in customer satisfaction as a result of increased performance inconsistency.

Next, we examine the situation where the firm does not collect any customer satisfaction data, a "what-if" scenario (cf. Wu et al. 2015). Thus, the firm has to predict the purchase probability conditional on interpurchase times using only objective service performance data. We first compute the hazard rate using only the (changed) objective service performance data and contrast it to the baseline case (no disconfirmation and no performance inconsistency). The results (Figure 2b) show that an increase of 50% in delivery time lengthens the interpurchase time by 2.3%. At the mean purchase amount per order of $23.34 in the data, this results in a revenue loss of approximately $7.2 million to the firm.

Finally, we pin down the incremental benefit of obtaining customer satisfaction data (in the presence of objective service performance data). We predict the the purchase probability using interpurchase times while excluding the residual from Equation 2 and compare it to the prediction with the residual as objective service performance increases (or decreases). Recall that the residual captures information over and above that in the objective service performance data. As can be seen from Figure 2c, when we take away the residual term from our proposed model, the predicted interpurchase time can be overstated up to 47.1%. Interestingly, as the change in delivery time gets larger and larger, the overstatement of the interpurchase time declines. This is because, for large changes, the objective service performance dominates the change in customer satisfaction, reducing the impact of the residual. A translation of the overstatement in interpurchase time into economic terms (at the mean order value) yields an amount of $207,520.40, a not inconsequential number.

Overall, this exercise illustrates that collecting customer satisfaction data, even in the presence of objective service performance data, makes economic and business sense for this firm. At the minimum, these data improve the firm's understanding of customer response, leading to a better calibration of business outcomes. Such an exercise also provides bounds on the amount that firms should be willing to spend to collect these data.

[Figure 2 about here.]

STUDY 2: AUTO RENTAL INDUSTRY

Institutional Background

In the auto rental industry, customer demand for different types (i.e., classes) of cars typically does not match the available inventory at rental locations (Carroll and Grimes 1995). This mismatch between demand and supply often results in free car-class upgrades, where customers receive a higher-class car for no extra charge (we describe the hierarchy of car-classes below). For example, when demand for a lower car-class exceeds the available inventory and the forecasted demand for higher car-class is low, auto rental companies often provide free upgrades, using non-utilized higher-class cars as a "cheap" way to avert customer complaints and increase customers' positive reactions (Hoffman, Kelley, and Chung

2003; Jiang, Hoegg, and Dahl 2013). Based on social exchange and equity theory (Walster, Berscheid, and Walster 1973), a service failure/recovery encounter can be viewed as an exchange in which the organization attempts to provide a gain, in the form of a recovery effort, to make up for the customer's loss and result in customer satisfaction (e.g., Smith, Bolton, and Wagner 1999; Knox and Van Oest 2014). Given the prevalence and frequency (54.6% in our data) of overbooking and free upgrades, however, we assume that customers perceive overbooking as a minor outcome failure[14] and as a result, consider a subsequent free car-class upgrade to be a better-than-expected service performance or a gain rather than a loss recovery.[15]

The auto rental firm tracks transaction-specific customer satisfaction from an online survey. In order to complete the survey, customers are provided with a hyperlink via email or on their printed receipt. Because the customers provide satisfaction data after each rental experience, we assume that their ratings reflect transaction-specific satisfaction with respect to the most recent service performance. The 10-point scale (except Q3: 5-point scale) satisfaction survey consists of the following eight questions:

[14] Services marketing literature recognizes two types of service encounter failures: outcome and process. The outcome dimension of a service encounter involves what customers actually receive from the service, whereas the process dimension involves how they receive the service, that is, the manner in which it is delivered (Smith, Bolton, and Wagner 1999). This paper focuses on outcome failures i.e., free upgrades.

[15] We verify our assessment (that a free car-class upgrade reflects better service performance) using data from an online survey. Specifically, we collected data from 450 U.S resident Mechanical Turk respondents in July 2016. We presented three service scenarios related to overbooking and recovery - car-class upgrade, hotel upgrade and car-class downgrade. We elicited responses to a variety of items, drawn from Maxham and Netemeyer (2002) and Smith, Bolton, and Wagner (1999), on a 7-point scale (a detailed list of survey questions and their ratings is available from the authors on request). We find that the mean service failure severity rating (7 being the severest service failure and 1 the weakest) are the lowest for car-class upgrade at 2.09 (the mean for car-class downgrade is 3.82 and that for hotel upgrade is 3.28), suggesting that car overbooking is not perceived by consumers as being a serious failure to be recovered from (perhaps due to its prevalence). Respondents also rated car upgrades as a better mechanism to resolve overbooking issues (mean = 5.33) than hotel upgrades (mean = 5.09) and car downgrades (mean = 3.71). Car upgrades were also seen as higher on distributive justice relative to hotel upgrades (mean = 5.57) and car downgrades (mean = 2.74). All means are different from each other at the p=0.01 level. These results provide evidence consistent with the belief that car upgrades are more likely to be perceived as better-than-expected service performance rather than as service failures.

- Q1: Please rate your overall experience.
- Q2: How likely is it that you would recommend Hertz to a friend or colleague?
- Q3: How likely are you to rent in the future?
- Please rate your experience with us in the following areas:
  - Q4: Courtesy of staff.
  - Q5: Speed of service.
  - Q6: Condition of vehicle & equipment.
  - Q7: Transaction and/or billing as expected.
  - Q8: Value for the money.

Data Description

The data used in this study come from a major car rental company in the United States. The data follow a panel of 454,597 unique loyalty club members from May 2010 through October 2012. The entire sample involves 2,981,503 rental car transactions across 3,422 locations (684 airport and 2,729 off-airport with 79.2% of the total transactions at airport locations) in the US.[16] Each location offers up to 23 different car classes, but five car classes - Compact (B), Intermediate (C), Standard (D), Full-Size (F), and Mid-Size SUV (L) - account for 90% of all transactions. Each record in the individual-level data corresponds to one rental and provides information on club membership ID,[17] store/rental location ID, the rental's check-out/in date, order number, pickup/return location, car class, rental price, price code (corporate/leisure), customer tier code and booking channel code. The data also record the car class reserved, the class received and the class charged for each customer

[16]From the original data that contain 6,283,105 observations we drop the transactions with invalid customer ID and missing car-class information. We also delete outliers ($> 99^{th}$ percentile) of rental duration, advance booking, rental price, and purchase frequency.

[17]We identify unique customers by a combination of membership IDs and birth dates on their driver's licenses. By doing this, we rule out the possibility that the purchase history under a single membership consists of multiple customers (i.e., drivers). 17.5% of the club members first appear in the data set on or after May 2010.

transaction. If the firm provided a higher car class than the one reserved and paid for, the transaction is considered a free upgrade - this occurs for 54.6% of all transactions.

Given that we need to use customer variation in objective service performance to create our measures (e.g., performance inconsistency), we focus on customers who purchased three or more times and provided satisfaction ratings at least once during the sample period. This results in 1,982,404 transactions from 126,246 customers. As can be seen from Table 9, except for the number of rentals, this sub-sample is very similar to the complete sample on behavioral metrics.

[Table 9 about here.]

From the eight questions on a 10-points scale in the online survey, we use Q4 (courtesy of staff), Q5 (speed of service), Q6 (condition of vehicle/equipment), Q7 (transaction/billing), and Q8 (value for money) to construct the overall measure of customer satisfaction. We do not use the first three questions because they are measures either of customer loyalty or akin to the Net Promoter Score, and thus conceptually distinct from customer satisfaction. Consistent with Study 1, we take the average of the ratings on these five questions to construct our customer satisfaction measure. This measure has a mean and median of 7.53 and 8.2, respectively. Figure 3 shows the distribution of the satisfaction measure and its correlation with free upgrade. A regression analysis of satisfaction ratings on free upgrades shows a positive and significant effect of free upgrades. On average, customers provides a satisfaction rating 1.34 times, or on 8.6% of all transactions.

[Figure 3 about here.]

Table 10 reports summary statistics on the key variables, including free upgrade, daily rental price, rental duration, and interpurchase time. We break up the data into the transactions with and without satisfaction ratings, in order to check if there is a systematic difference between these two groups of observations. We find that the average interpurchase time and the number of transactions since last rating to be substantially different across the two samples, with the former being longer and the latter being smaller relative to transactions with no rating. As in Study 1, these differences suggest the possibility of a within-individual selection problem. We thus use the transactions both with and without satisfaction rating to correct for selection. Note that customers were included in this dataset only if they were "raters," i.e., they participated in the survey at least once during the 24-month sample period. We therefore cannot examine across-individual selection in this Study.

[Table 10 about here.]

Results

The results from our proposed model (1) are reported in Table 11. In this table, we also include the results from a null model (2) where we do not correct for within-individual selection. In both specifications, unobserved customer heterogeneity is captured using the hierarchical Bayesian framework.

[Table 11 about here.]

As can be seen from the table, we find a significant effect of objective service performance on satisfaction rating. In particular, with greater disconfirmation (better-than-expected free upgrade offer) customers are more likely to provide higher satisfaction ratings. Furthermore, inconsistent service performance significantly decreases both survey participation and customer satisfaction. Second, the results show that both objective service performance and customer satisfaction (based on other factors than objective service performance) have a direct and positive effect on the probability of purchase conditional on interpurchase time. These results confirm that satisfaction ratings provide additional information over and above what can be obtained from objective service performance and that the two act as complements, not substitutes. Unlike the case of delivery time, the worst service performance customers can experience in the car-class upgrade context is to receive a car from the car-class that they originally reserved. Performance inconsistency thus captures fluctuation in performance only on the positive side. We speculate that this may drive performance inconsistency to have positive instead of negative effect on interpurchase time. We also confirm that within-individual selection needs to be addressed, that is, less satisfied customers are significantly more likely to rate. Note that ignoring this within-individual selection leads disconfirmation elasticity to customer satisfaction to be underestimated by 1.7% and performance inconsistency elasticity to be overestimated by 35.7%. Our instrument to identify the selection process, the number of transactions since the previous rating, significantly increases customers' participation in satisfaction rating. Overall, these results are consistent with those obtained from Study 1 - a quick comparison of the results and their consistency can be seen in Table 12.

[Table 12 about here.]

Given that our data in this study come from a single firm in a highly competitive industry, as a robustness check, we restrict our analysis to only corporate customers i.e., using only the transactions with "corporate" price codes in the data. Due to the nature of the corporate contracts, these customers are less likely to be affected by competition.[18] As shown in Table 11 (Column (3)), most of the results from these corporate customers are similar to those from the larger sample.[19] Notably, we replicate the finding that both direct and indirect (through customer satisfaction) effects of objective service performance are significant.

CONCLUSION

This paper examines two contemporaneous and growing trends in the service industry -

[18]We thank an anonymous reviewer for suggesting this analysis.

[19]Rental price does not have a direct impact on survey participation and purchase probability for these customers. This is probably due to fact that as customers are locked into a service provider (due to the corporate contract), price is not as important.

the collection of individual-level transaction-level customer satisfaction data and the availability of transaction-level objective service performance data. The research focus of the paper is to determine if these two data sources provide complementary or substitute insights to firms in terms of business outcomes. A potential issue with the analysis of such customer satisfaction data is within-individual selection, where the decision of which transaction to rate by an individual could be systematically related to the level of satisfaction. We leverage unique high-frequency individual-level transaction, satisfaction and objective service performance data from two different service industries to look at the relationship between these two types of data. We build a two-stage model where the explanatory variables are drawn from the theoretical literature on customer satisfaction. In the first stage, we examine each customer's decision to rate and conditional on that, the rating provided. The panel nature of our data allows us to control for within-individual selection in this stage. In the second stage, we fit consumers' interpurchase times using a proportional hazard model. To determine whether customer satisfaction and objective service performance data are complements or substitutes, we use the residuals from the first stage as a predictor of purchase behavior. These residuals represent the information available in the customer satisfaction ratings over and above the information provided in the objective service performance.

Our results suggest that these two sources of data are complementary with both the residual and (a function of) the objective performance data having a significant impact on the purchase outcomes. These findings are robust to a variety of assumptions regarding different independent and dependent variables as well as functional forms. In addition, we also find that there is a selection bias in the decision to rate. To the best of our knowledge, the finding that these two sources of data are complementary has not been documented before. Other results suggest that customer satisfaction is a function of both the level of the objective service performance and its consistency, leading to an asymmetric impact of increase versus decreasing objective service performance. We conduct a counterfactual analysis where firms do not have access to the customer satisfaction data (but do have access to the objective service performance data) and find that the prediction of business outcomes suffers in an economically meaningful manner. We are also able to provide a bound on the value of collecting customer satisfaction data. The fact that we find a similar pattern of results in the two very different service settings suggests that these results may be generalizable across multiple service industries.

Our research opens up avenues for further research, some driven by the limitations of our data. First, from a theoretical point of view, the source of the additional information in the customer satisfaction data needs to be investigated more carefully. It is possible that the satisfaction information may capture aspects of the transaction not reflected in the objective service data and/or may be conceptualized in the customer's mind as operating over a longer time period. Second, the availability of competitive data may allow us to develop boundary conditions for the relationship between the two sources of data. While we do offer a robustness check to competition in Study 2, we cannot provide an explicit control for competition given our data. Third, given that our data are secondary, we can control for self-selection only via the exclusion restriction. A perfect control for selection can only be implemented via an experimental procedure where customers are somehow assigned randomly to "rater" and "non-rater" conditions. Finally, the precision and ease of obtaining objective service performance measures may differ across service settings, leading to differences in the benefit of collecting customer satisfaction data. We hope that future research can implement our approach in many more settings to document these differences.

References

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

Bickart, Barbara, David Schmittlein. 1999. The Distribution of Survey Contact and Participation in the United States: Constructing a Survey-based Estimate. *Journal of Marketing Research* **36**(2) 286–294.

Bitner, Mary Jo. 1992. Servicescapes: the Impact of Physical Surroundings on Customers and Employees. *Journal of Marketing* **56**(2) 57–71.

Blattberg, Robert C, Scott A Neslin. 1990. *Sales Promotion: Concepts, Methods, and Strategies*. Prentice Hall.

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.

Bolton, Ruth N, Katherine N Lemon. 1999. A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction. *Journal of Marketing Research* **36**(2) 171–186.

Bolton, Ruth N., Katherine N. Lemon, Matthew D. Bramlett. 2006. The Effect of Service Experiences over Time on a Supplier's Retention of Business Customers. *Management Science* **52**(12) 1811–1823.

Bolton, Ruth N, Katherine N Lemon, Peter C Verhoef. 2008. Expanding Business-to-Business Customer Relationships: Modeling the Customer's Upgrade Decision. *Journal of marketing* **72**(1) 46–64.

Boulding, William, Ajay Kalra, Richard Staelin, Valarie A. Zeithaml. 1993. A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions. *Journal of Marketing Research* **30**(1) 7–27.

Carroll, William J., Richard C. Grimes. 1995. Evolutionary Change in Product Management: Experiences in the Car Rental Industry. *Interfaces* **25**(5) 84–104.

Chase, Richard B., Sriram Dasu. 2001. Want to Perfect Your Company's Service? Use Behavioral Science. *Harvard Business Review* **79**(6) 79–84.

Churchill, Gilbert A., Carol Surprenant. 1982. An Investigation into the Determinants of Customer Satisfaction. *Journal of Marketing Research* **19**(4) 491–504.

Cooil, Bruce, Timothy L Keiningham, Lerzan Aksoy, Michael Hsu. 2007. A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing* **71**(1) 67–83.

Couper, Mick P. 2000. Review: Web surveys: A Review of Issues and Approaches. *The Public Opinion Quarterly* **64**(4) 464–494.

Cox, David R. 1975. Partial likelihood. *Biometrika* **62**(2) 269–276.

Dong, Xiaojing, Ramkumar Janakiraman, Ying Xie. 2014. The Effect of Survey Participation on Consumer Behavior: The Moderating Role of Marketing Communication. *Marketing Science* **33**(4) 567–585.

Erdem, Tülin. 1998. An Empirical Analysis of Umbrella Branding. *Journal of Marketing Research* **35**(3) 339–351.

Fader, Peter S, James M Lattin, John DC Little. 1992. Estimating Nonlinear Parameters in the Multinomial Logit Model. *Marketing Science* **11**(4) 372–385.

Fishbein, Martin, Icek Ajzen. 1975. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA.

Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, Barbara Everitt Bryant. 1996. The American Customer Satisfaction Index: Nature, Purpose, and Findings. *Journal of Marketing* **60**(4) 7–18.

Fornell, Claes, Roland T Rust, Marnik G Dekimpe. 2010. The Effect of Customer Satisfaction on Consumer Spending Growth. *Journal of Marketing Research* **47**(1) 28–35.

Gijsenberg, Maarten J., Harald J. van Heerde, Peter C. Verhoef. 2015. Losses Loom Longer Than Gains: Modeling the Impact of Service Crises on Perceived Service Quality over Time. *Journal of Marketing Research* **17**(1) 642–656.

Godfrey, Andrea, Kathleen Seiders, Glenn B. Voss. 2011. Enough is Enough! The Fine Line in Executing Multichannel Relational Communication. *Journal of Marketing* **75**(4) 94–109.

Golder, Peter N, Gerard J Tellis. 2004. Growing, Growing, Gone: Cascades, Diffusion, and Turning Points in the Product Life Cycle. *Marketing Science* **23**(2) 207–218.

Grewal, Rajdeep, Murali Chandrashekaran, Alka V. Citrin. 2010. Customer Satisfaction Heterogeneity and Shareholder Value. *Journal of Marketing Research* **47**(4) 612–626.

Grimes, William. 2012. When Businesses Can't Stop Asking, 'How Am I Doing?'. *The New York Times* .

Groves, Robert M., Robert B. Cialdini, Mick P. Couper. 1992. Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly* **56**(4) 475–495.

Gupta, Sunil. 1991. Stochastic Models of Interpurchase Time with Time-Dependent Covariates. *Journal of Marketing Research* **28**(1) 1–15.

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* **47**(1) 153–161.

Helsen, Kristiaan, David C Schmittlein. 1993. Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models. *Marketing Science* **12**(4) 395–414.

Hess, Ronald L., Shankar Ganesan, Noreen M. Klein. 2003. Service Failure and Recovery: The Impact of Relationship Factors on Customer Satisfaction. *Journal of the Academy of Marketing Science* **31**(2) 127–145.

Hoffman, K. Douglas, Scott W. Kelley, Beth C. Chung. 2003. A CIT Investigation of Servicescape Failures and Associated Recovery Strategies. *Journal of Services Marketing* **17**(4) 322–340.

Jiang, Lan, JoAndrea Hoegg, Darren W Dahl. 2013. Consumer Reaction to Unearned Preferential Treatment. *Journal of Consumer Research* **40**(3) 412–427.

Johnson, Michael D., Eugene W. Anderson, Claes Fornell. 1995. Rational and Adaptive Performance Expectations in a Customer Satisfaction Framework. *Journal of Consumer Research* **21**(4) 695–707.

Kahneman, Daniel, Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47**(2) 263–291.

Kai, Li. 1998. Bayesian Inference in a Simultaneous Equation Model with Limited Dependent Variables. *Journal of Econometrics* **85**(2) 387–400.

Keiningham, Timothy L., Lerzan Aksoy, Edward C. Malthouse, Bart Lariviere, Alexander Buoye. 2014a. The Cumulative Effect of Satisfaction with Discrete Transactions on Share of Wallet. *Journal of Service Management* **25**(3) 310–333.

Keiningham, Timothy L, Forrest V Morgeson, Lerzan Aksoy, Luke Williams. 2014b. Service Failure Severity, Customer Satisfaction, and Market Share an Examination of the Airline Industry. *Journal of Service Research* **17**(4) 415–431.

Knox, George, Rutger Van Oest. 2014. Customer Complaints and Recovery Effectiveness: A Customer Base Approach. *Journal of Marketing* **78**(5) 42–57.

Lund, Donald J., Detelina Marinova. 2014. Managing Revenue Across Retail Channels: The Interplay of Service Performance and Direct Marketing. *Journal of Marketing* **78**(5) 99–118.

Luo, Xueming, Christian Homburg, Jan Wieseke. 2010. Customer Satisfaction, Analyst Stock Recommendations, and Firm Value. *Journal of Marketing Research* **47**(6) 1041–1058.

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, Pradeep K Chintagunta. 2006. The Effect of Banner Advertising on Internet Purchasing. *Journal of Marketing Research* **43**(1) 98–108.

Maxham, James G, III, Richard G Netemeyer. 2002. A Longitudinal Study of Complaining Customers? Evaluations of Multiple Service Failures and Recovery Efforts. *Journal of Marketing* **66**(4) 57–71.

McCollough, Michael A., Leonard L. Berry, Manjit S. Yadav. 2000. An Empirical Investigation of Customer Satisfaction After Service Failure and Recovery. *Journal of Service Research* **3**(2) 121–137.

Mithas, Sunil, Mayuram S. Krishnan, Claes Fornell. 2005. Why Do Customer Relationship Management Applications Affect Customer Satisfaction? *Journal of Marketing* **69**(4) 201–209.

Mitra, Debanjan, Peter N. Golder. 2006. How Does Objective Quality Affect Perceived Quality? Short-Term Effects, Long-Term Effects, and Asymmetries. *Marketing Science* **25**(3) 230–247.

Mittal, Vikas, Wagner A. Kamakura. 2001. Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing Research* **38**(1) 131–142.

Morgan, Neil A, Lopo Leotte Rego. 2006. The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance. *Marketing Science* **25**(5) 426–439.

Narayanan, Sridhar, Puneet Manchanda. 2012. An Empirical Analysis of Individual Level Casino Gambling Behavior. *Quantitative Marketing and Economics* **10**(1) 27–62.

Nerlove, Marc. 1958. Adaptive Expectations and Cobweb Phenomena. *The Quarterly Journal of Economics* **72**(2) 227–240.

Ofir, Chezy, Itamar Simonson. 2007. The Effect of Stating Expectations on Customer Satisfaction and Shopping Experience. *Journal of Marketing Research* **44**(1) 164–174.

Oliver, Richard L. 1977. Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation. *Journal of Applied Psychology* **62**(4) 480–486.

Oliver, Richard L. 1980. A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research* **17**(4) 460–469.

Olsen, Line Lervik, Michael D. Johnson. 2003. Service Equity, Satisfaction, and Loyalty: From Transaction-Specific to Cumulative Evaluations. *Journal of Service Research* **5**(3) 184–195.

Parasuraman, A., Valarie A. Zeithaml, Leonard L. Berry. 1988. SERVQUAL: A Multiple-Item Scale For Measuring Consumer Perception of Service Quality. *Journal of Retailing* **64**(1) 12–40.

Petrin, Amil, Kenneth Train. 2010. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* **47**(1) 3–13.

Rust, Roland T., J. Jeffrey Inman, Jianmin Jia, Anthony J. Zahorik. 1999. What You Don't Know about Customer-Perceived Quality: The Role of Customer Expectation Distributions. *Marketing Science* **18**(1) 77–92.

Rust, Roland T, Katherine N Lemon, Valarie A Zeithaml. 2004. Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing* **68**(1) 109–127.

Rust, Roland T, Anthony J Zahorik. 1993. Customer Satisfaction, Customer Retention, and Market Share. *Journal of Retailing* **69**(2) 193–215.

Seiders, Kathleen, Glenn B. Voss, Dhruv Grewal, Andrea L. Godfrey. 2005. Do Satisfied Customers Buy More? Examining Moderating Influences in a Retailing Context. *Journal of Marketing* **69**(4) 26–43.

Smith, Amy K, Ruth N Bolton, Janet Wagner. 1999. A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery. *Journal of Marketing Research* **36**(3) 356–372.

Sriram, S, Pradeep K Chintagunta, Puneet Manchanda. 2015. Service Quality Variability and Termination Behavior. *Management Science* **61**(11) 2739–2759.

Van Doorn, Jenny, Peter C Verhoef. 2008. Critical Incidents and the Impact of Satisfaction on Customer Share. *Journal of Marketing* **72**(4) 123–142.

Verhoef, Peter C. 2003. Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development. *Journal of marketing* **67**(4) 30–45.

Verma, Rohit, Gary M. Thompson, Jordan J. Louviere. 1999. Configuring Service Operations in Accordance With Customer Needs and Preferences. *Journal of Service Research* **1**(3) 262–274.

Voss, Glenn B, Andrea Godfrey, Kathleen Seiders. 2010. How Complementarity and Substitution Alter the Customer Satisfaction-Repurchase Link. *Journal of Marketing* **74**(6) 111–127.

Voss, Glenn B, Ananthanarayanan Parasuraman, Dhruv Grewal. 1998. The Roles of Price, Performance, and Expectations in Determining Satisfaction in Service Exchanges. *Journal of Marketing* **62**(4) 46–61.

Walster, Elaine, Ellen Berscheid, G William Walster. 1973. New Directions in Equity Research. *Journal of Personality and Social Psychology* **25**(2) 151–176.

Weaver, Donald, Philip Brickman. 1974. Expectancy, Feedback, and Disconfirmation as Independent Factors in Outcome Satisfaction. *Journal of Personality and Social Psychology* **30**(3) 420–428.

Wu, Chunhua, Hai Che, Tat Y Chan, Xianghua Lu. 2015. The Economic Value of Online Reviews. *Marketing Science* **34**(5) 739–754.

Zhang, Wei, Ajay Kalra. 2014. Joint Examination of Quality Choice and Satisfaction: The Impact of Circumstantial Variables. *Journal of Marketing Research* **51**(4) 448–462.

Table 1: Prior Research on the Longitudinal Analysis of Customer Satisfaction and Business Outcomes

Study	Research Context	Dependent Variable(s)	Satisfaction Measure	Study Period	Self-Selection Correction
Bolton (1998)	Telecommunication	Duration of Relationship	Single Cumulative Rating	1 year	X
Bolton and Lemon (1999)	TV/Telecommunication	Minutes of Use per Month	Single Cumulative Rating	6 months	X
Bolton et al. (2006)	System Support Service Contracts	Contract Renewal	-	3 years	-
Bolton et al. (2008)	System Support Service Contracts	Upgraded Contracts	Single Cumulative Rating	3 years	X
Mittal and Kamakura (2001)	Automotive Service	Repurchase Behavior	Single Cumulative Rating	Varies	X
Seiders et al. (2005)	Fashion Apparel/Home Furnishing	Repurchase Visits/Spending	Single Cumulative Rating	1 year	across-individual
Voss et al. (2010)	Fashion Apparel/Automotive Service	Repurchase Visits/Spending	Single Cumulative Rating	1 year	across-individual
Cooil et al. (2007)	Canadian Baking Service	Customer Share of Wallet	Multiple Cumulative Ratings	5 years	X
Keiningham et al. (2014a)	Specialty Retailer	Customer Share of Wallet	Multiple Transaction-Specific Ratings	21 months	X
Van Doorn and Verhoef (2008)	Logistics Service	Customer Share of Wallet	Multiple Cumulative Ratings	3 years	X
Fornell et al. (2010)	43 Different Industries	Consumer Spending	Multiple Cumulative Ratings (ACSI)	12 years	X
Luo et al. (2010)	24 Different Industries	Stock Recommendations	Multiple Cumulative Ratings (ACSI)	12 years	X
Grewal et al. (2010)	Domestic Airline Service	Shareholder Value	Multiple Cumulative Ratings (ACSI)	9 years	X
Current Study	Quick Service Restaurant/Auto Rental	Interpurchase Time	Multiple Transaction-Specific Ratings	1/2 year(s)	Within-individual

Table 2: Descriptive Statistics (Study 1)

(a) Transactions from raters (n = 632,600)

	Mean	SD	Median	Min.	Max.	Skewness	$\mathrm{Cor}_{t,t-1}$
Delivery Time (minutes)	36.9	13.1	35	3	119	0.78	0.24
Purchase Amount (dollar)	23.24	10.03	21.11	0.01	828.5	4.7	0.45
Coupon Redemption	0.51	0.5	1	0	1	-0.03	0.43
Interpurchase Time (day)	29.81	37.24	15	1	345	2.69	0.15
Number of Transactions Since Rating_{t-1}	4.57	4.91	3	1	106	4.48	0.9

(b) Transactions from raters with 3 or more purchases (n = 484,440)

	Mean	SD	Median	Min.	Max.	Skewness	$\mathrm{Cor}_{t,t-1}$
Delivery Time (minutes)	36.84	12.83	35	3	119	0.81	0.24
Purchase Amount (dollar)	23.34	10.09	21.24	0.01	828.5	4.9	0.45
Coupon Redemption	0.5	0.5	1	0	1	-0.02	0.44
Interpurchase Time (day)	28.21	35.27	15	1	345	2.73	0.15
Number of Transactions Since Rating_{t-1}	4.98	5.15	4	1	106	4.26	0.9

Table 3: Descriptive Statistics: Within-Individual Selection (Study 1)

(a) Transactions from raters with 3 or more purchases: with rating (n = 116,096)

	Mean	SD	Median	Min.	Max.	Skewness	$\mathrm{Cor}_{t,t-1}$
Delivery Time (minutes)	37.44	12.17	35	3	119	1.29	0.22
Purchase Amount (dollar)	23.34	9.42	21.39	0.06	361.8	2.83	0.48
Coupon Redemption	0.55	0.5	1	0	1	-0.21	0.42
Interpurchase Time (day)	38.73	46.13	21	1	345	2.24	0.09
Number of Transactions Since Rating_{t-1}	3.96	3.8	3	1	106	3.9	0.89

(b) Transactions from raters with 3 or more purchases: without rating (n = 368,344)

	Mean	SD	Median	Min.	Max.	Skewness	$\mathrm{Cor}_{t,t-1}$
Delivery Time (minutes)	36.65	13.03	35	3	119	0.69	0.25
Purchase Amount (dollar)	23.34	10.3	21.19	0.01	828.5	5.39	0.44
Coupon Redemption	0.49	0.5	0	0	1	0.04	0.44
Interpurchase Time (day)	24.89	30.32	14	1	335	2.72	0.15
Number of Transactions Since Rating_{t-1}	5.3	5.47	4	1	105	4.18	0.9

Table 4: Descriptive Statistics: across-individual Selection (Study 1)

(a) Transactions from raters (n = 632,600)

	Mean	SD	Median	Min.	Max.	Skewness	$Cor_{t,t-1}$
Delivery Time (minutes)	36.9	13.1	35	3	119	0.78	0.24
Purchase Amount (dollar)	23.24	10.03	21.11	0.01	828.5	4.7	0.45
Coupon Redemption	0.51	0.5	1	0	1	-0.03	0.43
Interpurchase Time (day)	29.81	37.24	15	1	345	2.69	0.15
Number of Transactions Since $Rating_{t-1}$	4.57	4.91	3	1	106	4.48	0.9

(b) Transactions from non-raters (n = 6,655,320)

	Mean	SD	Median	Min.	Max.	Skewness	$Cor_{t,t-1}$
Delivery Time (minutes)	35.38	14.66	35	3	119	0.41	0.33
Purchase Amount (dollar)	22.47	12.5	20.11	0.01	978.9	11.19	0.49
Coupon Redemption	0.36	0.48	0	0	1	0.59	0.39
Interpurchase Time (day)	37.25	45.83	20	1	358	2.36	0.13

Table 5: Parameter Estimates from the Selection Model (Study 1)

	Proposed Model (1)			Without Selection (2)		
	Mean	2.5%CI	97.5%CI	Mean	2.5%CI	97.5%CI
Rating Incidence Model						
Intercept	-0.6775	-0.6918	-0.6628			
Disconfirmation	0.2382	0.2187	0.2578			
Performance Inconsistency	-0.4305	-0.4897	-0.3761			
Coupon Redemption	0.1738	0.1636	0.1842			
Number of Transactions Since Rating_{t-1}	0.0045	0.0020	0.0071			
Satisfaction Rating Model						
Intercept	4.7951	4.7796	4.8092	4.7456	4.7361	4.7549
Disconfirmation	-0.3801	-0.3994	-0.3614	-0.3734	-0.3925	-0.3534
Performance Inconsistency	-0.4568	-0.5040	-0.4123	-0.4771	-0.5247	-0.4315
Coupon Redemption	0.0118	0.0042	0.0200	0.0168	0.0095	0.0246
Inverse Mills Ratio	-0.0406	-0.0498	-0.0305			
Smoothing Factor (δ)	0.3			0.3		
DIC	321897			352908		
Number of Observations	116096			116096		

* The estimates of the hyperparameters (that capture unobserved heterogeneity) are omitted in the interest of brevity.

Table 6: Parameter Estimates from the Probability of Purchase Model (Study 1)

	Proposed Model			Partial Models											
				Actual Satisfaction Rating			Actual Service Measure			Constructed Service Measures			Actual Satisfaction Rating & Actual Service Measure		
	(1)			(2)			(3)			(4)			(5)		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Baseline Hazard (h_0)	-0.4987	-0.6429	-0.3734	-0.6134	-0.7853	-0.4474	-0.4253	-0.5723	-0.3072	-0.4483	-0.5963	-0.3018	-0.4025	-0.5715	-0.2544
Residuals	0.0683	0.0558	0.0829										0.0734	0.0582	0.0867
Actual Satisfaction				0.0863	0.0777	0.0948									
Actual Delivery Time							-0.0374	-0.0663	-0.0102				-0.0514	-0.0803	-0.0208
Disconfirmation	0.0098	-0.0172	0.0369							-0.0021	-0.0282	0.0235			
Performance Inconsistency	-0.0759	-0.1335	-0.0106							-0.0914	-0.1467	-0.0295			
Coupon Redemption	-0.0791	-0.0913	-0.0672	-0.0920	-0.1040	-0.0804	-0.0840	-0.0958	-0.0725	-0.0841	-0.0952	-0.0724	-0.0803	-0.0921	-0.0696
Purchase Amount	0.0119	-0.0056	0.0278	-0.0007	-0.0178	0.0159	0.0147	-0.0018	0.0312	0.0173	0.0017	0.0348	0.0097	-0.0079	0.0278
lag (Interpurchase Time)	-0.1618	-0.1672	-0.1564	-0.1658	-0.1723	-0.1590	-0.1593	-0.1651	-0.1531	-0.1612	-0.1669	-0.1553	-0.1605	-0.1664	-0.1546
RMSE		0.4872			0.6424			0.5278			0.5244			0.4950	
MAE		0.4112			0.4783			0.4441			0.4401			0.4156	
DIC		2433955			2421806			2432996			2433849			2432929	
Number of Observations		116096			116096			116096			116096			116096	

* The estimates of the hyperparameters (that capture unobserved heterogeneity) are omitted in the interest of brevity.

Table 7: Parameter Estimates from the Robustness Checks (Study 1)

	Purchase Amount (1)			Expectation: Recent (2)			Asymmetry of Disconfirmation (3)		
	Mean	2.5%CI	97.5%CI	Mean	2.5%CI	97.5%CI	Mean	2.5%CI	97.5%CI
Rating Incidence Model									
Intercept	-0.6775	-0.6918	-0.6628	-0.6692	-0.6832	-0.6554	-0.6726	-0.6869	-0.6569
Disconfirmation	0.2382	0.2187	0.2578	0.1735	0.1573	0.1908			
Negative Disconfirmation							0.1719	0.1383	0.2048
Positive Disconfirmation							0.3380	0.2960	0.3823
Performance Inconsistency	-0.4305	-0.4897	-0.3761	-0.4757	-0.5351	-0.4176	-0.3146	-0.3816	-0.2513
Coupon Redemption	0.1738	0.1636	0.1842	0.1730	0.1630	0.1830	0.1704	0.1606	0.1798
Number of Transactions Since Rating_{t-1}	0.0045	0.0020	0.0071	0.0031	0.0009	0.0055	0.0034	0.0010	0.0062
Satisfaction Rating Model									
Intercept	4.7951	4.7796	4.8092	4.8007	4.7852	4.8158	4.8104	4.7962	4.8228
Disconfirmation	-0.3801	-0.3994	-0.3614	-0.2504	-0.2650	-0.2347			
Negative Disconfirmation							-0.7367	-0.7665	-0.7060
Positive Disconfirmation							0.0500	0.0107	0.0867
Performance Inconsistency	-0.4568	-0.5040	-0.4123	-0.5199	-0.5701	-0.4694	-0.1643	-0.2144	-0.1236
Coupon Redemption	0.0118	0.0042	0.0200	0.0127	0.0051	0.0204	0.0117	0.0045	0.0199
Inverse Mills Ratio	-0.0406	-0.0498	-0.0305	-0.0392	-0.0493	-0.0291	-0.0476	-0.0582	-0.0372
Probability of Purchase Model									
Intercept	1.2332	1.2160	1.2508						
Baseline Hazard (h_0)				-0.4793	-0.6175	-0.3384	-0.4180	-0.5554	-0.3037
Residuals	0.0079	0.0052	0.0106	0.0698	0.0555	0.0825	0.0641	0.0503	0.0771
Disconfirmation	-0.0093	-0.0172	-0.0014	-0.0342	-0.0582	-0.0119			
Performance Inconsistency	-0.0150	-0.0338	0.0035	-0.0849	-0.1414	-0.0254			
Actual Delivery Time							-0.0393	-0.0680	-0.0119
Coupon Redemption	0.0012	-0.0021	0.0045	-0.0835	-0.0958	-0.0708	-0.0884	-0.1023	-0.0755
Purchase Amount	0.5989	0.5936	0.6041	0.0061	-0.0112	0.0229	0.0177	0.0012	0.0344
lag (Interpurchase Time)	0.0013	-0.0001	0.0027	-0.1580	-0.1641	-0.1518	-0.1738	-0.1805	-0.1675
DIC		329526			2434691			2432638	
Number of Observations		116096			116096			116096	

* The estimates of the hyperparameters (that capture unobserved heterogeneity) are omitted in the interest of brevity.

Table 8: Parameter Estimates from the Robustness Checks (Study 1, Continued)

	Proposed Model (1)			Across-Individual Selection (2)		
	Mean	2.5%CI	97.5%CI	Mean	2.5%CI	97.5%CI
Rating Incidence Model						
Intercept	-0.6775	-0.6918	-0.6628	-1.9496	-2.0048	-1.8946
Disconfirmation	0.2382	0.2187	0.2578	0.1910	0.1126	0.2776
Performance Inconsistency	-0.4305	-0.4897	-0.3761	-0.0554	-0.2245	0.1063
Coupon Redemption	0.1738	0.1636	0.1842	0.3307	0.2866	0.3798
Number of Transactions Since Rating$_{t-1}$	0.0045	0.0020	0.0071	-0.0746	-0.0843	-0.0662
Satisfaction Rating Model						
Intercept	4.7951	4.7796	4.8092	5.0488	4.3236	5.7637
Disconfirmation	-0.3801	-0.3994	-0.3614	-0.3154	-0.5498	-0.0964
Performance Inconsistency	-0.4568	-0.5040	-0.4123	-0.4707	-0.9312	0.0340
Coupon Redemption	0.0118	0.0042	0.0200	-0.0739	-0.2081	0.0603
Inverse Mills Ratio	-0.0406	-0.0498	-0.0305	-0.0811	-0.3594	0.1991
Smoothing Factor (δ)		0.3			0.3	
DIC		321897			4397	
Number of Observations		116096			74080	

* The estimates of the hyperparameters (that capture unobserved heterogeneity) are omitted in the interest of brevity.

Table 9: Descriptive Statistics (Study 2)

(a) Transactions from all raters (n = 2,981,503)

	Mean	SD	Median	Min.	Max.	Skewness	$\text{Cor}_{t,t-1}$
Upgrade Probability	0.55	0.5	1	0	1	-0.2	0.13
Purchase Amount (dollar)	34.9	18.58	31.5	0	120	1.21	0.34
Rental Duration (day)	4.23	2.94	4	1	29	2.95	0.37
Interpurchase Time (day)	49.65	76.8	21	1	802	3.41	0.26
Number of Transactions Since Rating_{t-1}	8.63	10.01	5	1	152	2.9	0.96

(b) Transactions from raters with 3 or more purchases (n = 1,982,404)

	Mean	SD	Median	Min.	Max.	Skewness	$\text{Cor}_{t,t-1}$
Upgrade Probability	0.58	0.49	1	0	1	-0.3	0.13
Purchase Amount (dollar)	34.42	18.16	31.49	0	120	1.15	0.34
Rental Duration (day)	4.06	2.63	4	1	29	2.9	0.36
Interpurchase Time (day)	39.35	58.64	18	1	764	3.6	0.26
Number of Transactions Since Rating_{t-1}	8.97	10.01	5	1	152	2.83	0.95

Table 10: Descriptive Statistics: Within-Individual Selection (Study 2)

(a) Transactions from raters with 3 or more purchases: with rating (n = 87,329)

	Mean	SD	Median	Min.	Max.	Skewness	$\text{Cor}_{t,t-1}$
Upgrade Probability	0.55	0.5	1	0	1	-0.19	0.1
Purchase Amount (dollar)	33.57	18.82	30	0	120	1.27	0.32
Rental Duration (day)	4.47	3.14	4	1	29	2.63	0.35
Interpurchase Time (day)	77.63	92.09	42	1	762	2.19	0.19
Number of Transactions Since Rating_{t-1}	6.99	7.45	4	1	136	3.57	0.98

(b) Transactions from raters with 3 or more purchases: without rating (n = 1,895,075)

	Mean	SD	Median	Min.	Max.	Skewness	$\text{Cor}_{t,t-1}$
Upgrade Probability	0.58	0.49	1	0	1	-0.32	0.13
Purchase Amount (dollar)	34.51	18.08	31.58	0	120	1.14	0.34
Rental Duration (day)	4.01	2.57	4	1	29	2.92	0.36
Interpurchase Time (day)	35.26	52.23	16	1	764	3.82	0.25
Number of Transactions Since Rating_{t-1}	9.18	10.22	6	1	152	2.76	0.95

Table 11: Parameter Estimates from the Selection Model (Study 2)

	Proposed Model (1)			Without Selection (2)			Business Customers (3)		
	Mean	2.5%CI	97.5%CI	Mean	2.5%CI	97.5%CI	Mean	2.5%CI	97.5%CI
Rating Incidence Model									
Intercept	-1.0810	-1.1037	-1.0604				-1.3241	-1.4203	-1.2348
Disconfirmation	-0.0004	-0.0085	0.0075				-0.0076	-0.0360	0.0193
Performance Inconsistency	-0.2830	-0.3218	-0.2429				-0.4030	-0.5373	-0.2611
Rental Price	-0.0013	-0.0015	-0.0010				0.0002	-0.0015	0.0019
Number of Transactions Since Rating$_{t-1}$	0.0363	0.0346	0.0382				0.1119	0.1021	0.1218
Satisfaction Rating Model									
Intercept	7.9193	7.8634	7.9758	7.9092	7.8698	7.9478	7.9455	7.7762	8.1081
Disconfirmation	0.0949	0.0740	0.1160	0.0933	0.0730	0.1142	0.1652	0.1023	0.2256
Performance Inconsistency	-0.0942	-0.1586	-0.0261	-0.1279	-0.1904	-0.0593	-0.0261	-0.1993	0.1535
Rental Price	-0.0089	-0.0096	-0.0082	-0.0089	-0.0095	-0.0081	-0.0135	-0.0171	-0.0100
Inverse Mills Ratio	-0.0342	-0.0628	-0.0041				-0.0074	-0.0666	0.0544
Probability of Purchase Model									
Baseline Hazard (h_0)	2.0538	2.0311	2.0777	2.0455	2.0220	2.0700	5.8959	4.6780	6.7156
Residual	0.0164	0.0112	0.0214	0.0174	0.0129	0.0222	0.0255	0.0100	0.0434
Disconfirmation	0.0258	0.0140	0.0376	0.0258	0.0141	0.0379	0.0778	0.0410	0.1169
Performance Inconsistency	0.1151	0.0814	0.1486	0.1155	0.0820	0.1511	0.1349	0.0245	0.2360
Rental Price	-0.0017	-0.0020	-0.0013	-0.0017	-0.0020	-0.0014	0.0012	-0.0018	0.0046
lag (Interpurchase Time)	-0.1773	-0.1825	-0.1720	-0.1774	-0.1824	-0.1720	-0.2529	-0.2779	-0.2248
Smoothing Factor (δ)		0.3			0.3			0.3	
DIC		1809255			1809233			154938	
Number of Observations		87329			87329			10848	

* The estimates of the hyperparameters (that capture unobserved heterogeneity) are omitted in the interest of brevity.

Table 12: A Comparison of Results across the Two Different Service Settings

(a) Quick Service Restaurant Industry

	Rating Incidence	Satisfaction Rating	Purchase probability
Disconfirmation	+	−	
Performance Inconsistency	−	−	−
Customer Satisfaction			+

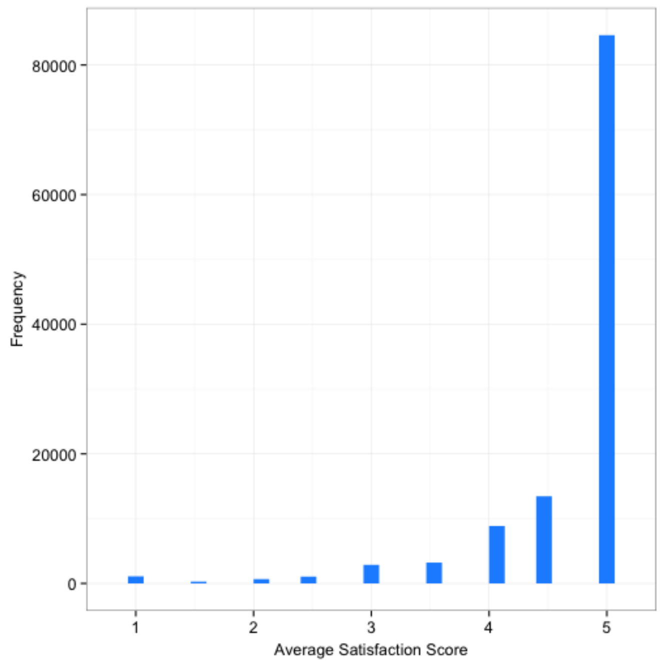
(b) Auto Rental Industry

	Rating Incidence	Satisfaction Rating	Purchase Probability
Disconfirmation*		+	+
Performance Inconsistency	−	−	+
Customer Satisfaction			+

* Disconfirmation of free upgrades expectation implies the opposite direction to disconfirmation of delivery time in (a).

Figure 1: Summary of Satisfaction Score (Study 1)

(a) Distribution of Satisfaction Score



(b) Relationship with Delivery Time

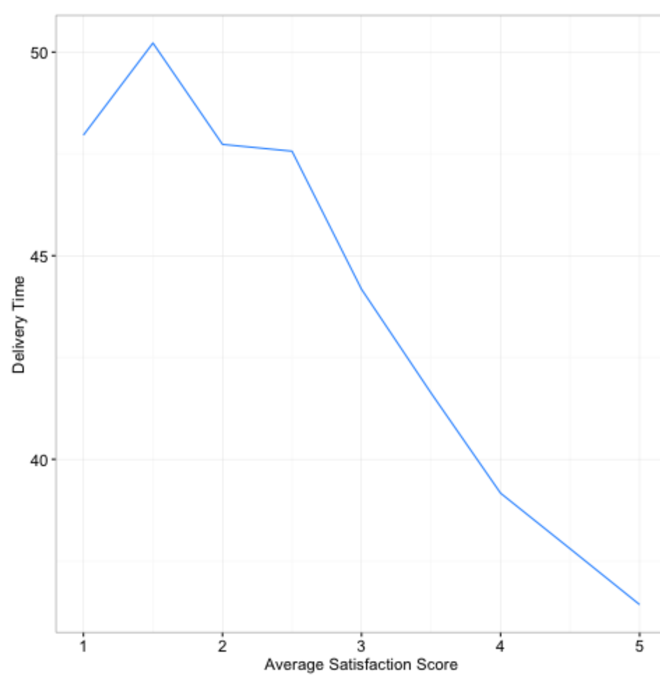
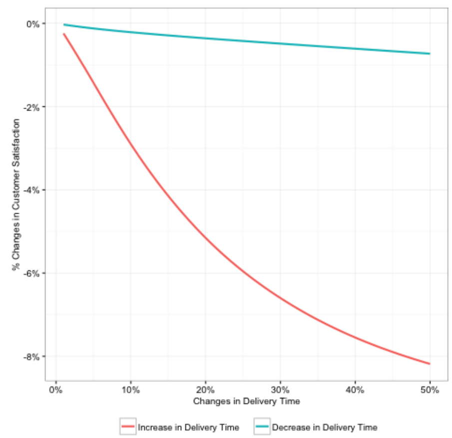
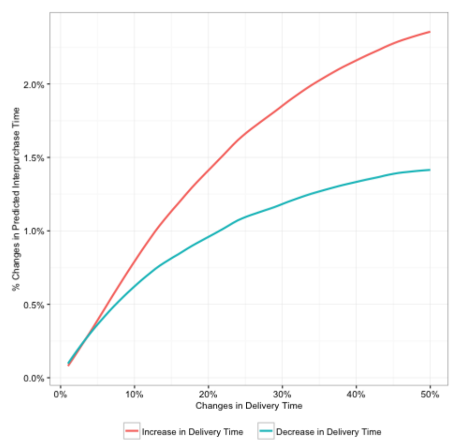


Figure 2: The Effect of Changes in Service

(a) Percent Changes in Predicted Customer Satisfaction Score



(b) Percent Changes in Predicted Interpurchase Time



(c) Contribution of the Residuals to Predicted Interpurchase Time

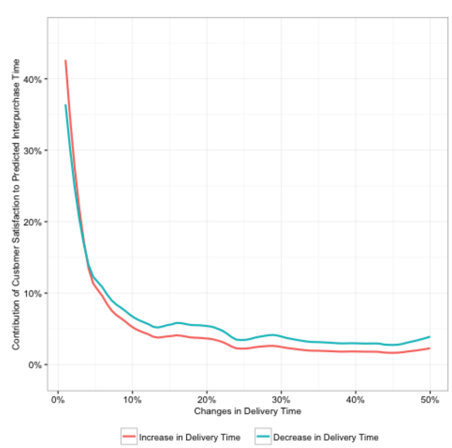


Figure 3: Summary of Satisfaction Score (Study 2)