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
October 2015

The Value of Measuring Customer Satisfaction

Jihoon Cho, Anocha Aribarg, and Puneet Manchanda

July 2015
This version: October 2015

Abstract

Firms in multiple industries collect customer satisfaction data to measure service performance. Increasingly, however, many firms are now able to collect objective service performance data as well. This raises the question of whether measuring customer satisfaction is valuable to firms when objective service performance data are available. The authors answer 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 repurchase behavior. The data come from two different - quick service restaurant and auto rental - service industries. The authors find that satisfaction ratings reflect objective service performance. Interestingly, despite its weak direct effect, the authors find a strong indirect effect of objective service performance on interpurchase time, operating through customer satisfaction. Unlike past research, these results are obtained after controlling for within-customer selection (of service encounters). Overall, the results suggest that customer satisfaction ratings are valuable.

Keywords: *Customer Satisfaction, Service Quality, Expectation Disconfirmation, Selection Bias, Performance Inconsistency*

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 two anonymous companies and the Wharton Customer Analytics Initiative for providing one of the datasets (auto rental) used in this paper. 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 have invested significantly in measuring and tracking customer satisfaction. As of 2012, spending on customer satisfaction research exceeded $750 million in the United States (Inside Research 2012). Firms use customer satisfaction as a measure of their service performance since objective measures of service performance are usually hard to obtain. For example, in a survey of nearly 200 senior marketing managers, 71 percent responded that they found the customer satisfaction metric very useful in managing and monitoring their businesses (Farris et al. 2010). Anecdotal evidence suggests that higher customer satisfaction does indeed lead to better firm outcomes e.g., the American Express 2014 Global Customer Service Barometer finds three out of four customers say they have spent more with a company because of a history of satisfactory customer service experiences (Ebiquity 2014). Conversely, bad customer service costs organizations in the United States $229 on average per year for each lost business relationship, making every single customer worth fighting for (Urlocker 2012).

Although customer satisfaction plays a key role in driving desirable customer behavior, firms find it difficult to obtain such (attitudinal) measures for every transaction in a cost effective manner. With more advanced information technology, many firms in service settings currently have capability to monitor objective service performance - mostly time based- at lower costs. For example, airlines track the percentage of on-time flights (Grewal et al. 2010), UPS uses real-time delivery tracking (Lund and Marinova 2014), and McDonald's monitors its drive-through service time (Hess et al. 2003). Given this, an emerging question is whether firms can bypass the collection of customer satisfaction data and rely solely on objective service performance data instead. For this to be possible, it is important to verify that customer satisfaction measures do indeed reflect objective service quality and they do not provide any additional information (when objective service quality measures are available). Thus, in this research, we aim to assess the value of measuring customer satisfaction by asking two specific questions - (a) do customer satisfaction measures reflect firm's objective service performance and (b) do customer satisfaction measures provide additional information relative to what can be obtained from objective service performance?

The academic literature (e.g., Oliver 1980) views customer satisfaction as being driven by the discrepancy between customer expectation and service performance - labeled "disconfirmation." Facing the same constraints as firms, most academic studies also rely on survey based self-reported measures of perceived service performance, customer satisfaction and purchase intent (a few no-table exceptions are discussed in a subsequent section). This prevalent method suffers from three limitations. First, in contrast to objective performance, perceived performance is derived from customers' subjective judgment of the observed performance. A variety of factors such as marketing communications and customers' prior expectation of performance could contribute to the discrepancy between perceived and observed service performance. Despite the customer-centric view that only perceived service performance matters to firms (e.g., Oliver 1980), understanding how objective service performance is linked to customers' behavior can provide firms with more actionable guidelines for improvement (e.g., Grewal et al. 2010). Second, not all customers reply to satisfaction surveys, opening up the possibility of the results being biased by self-selection. Failing to account for this self-selection will lead to biased inferences with regard to the observable factors that drive satisfaction ratings. Lastly, examining variables that are collected from the same survey is prone to high common-methods variance and likely inflates the relationships among the constructs under investigation (Fishbein and Ajzen 1975).

Unlike most previous customer satisfaction research that uses cross-sectional self-report survey data, our data consist of individual-level cross-sectional and time-series measures of objective service performance, customer satisfaction and repurchase behavior. Our data come from two very different - quick service restaurant and auto rental - service industries. Observing objective performance helps us validate the use of customer satisfaction as a proxy for service performance, and more importantly, allows us to examine the value of customer satisfaction when it is feasible for firms to measure objective service performance. The availability of transactions with and without satisfaction ratings allows us to correct for the presence of within-individual selection bias, something previous research has ignored. Therefore, our paper's main contribution is to examine the interplay among objective service performance, customer satisfaction, and actual purchase behavior while properly correcting for within-individual self-selection (in customer satisfaction ratings).

We achieve our research goal by modeling 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 and the second equation capturing the drivers of the satisfaction rating. In the second stage, we model customers' interpurchase time as a function of the predicted customer satisfaction rating (along with control variables). We incorporate random effects for the key parameters to allow for heterogeneous customer responses.

We find that customer satisfaction is correlated with objective service performance. Specifically, customer satisfaction is influenced by disconfirmation and service performance inconsistency. Interestingly, despite the absence of direct effects of disconfirmation and performance inconsistency, customer satisfaction has a strong effect on interpurchase time. We obtain these results after correcting for within-individual self-selection (we find that the decision to rate and the satisfaction rating are correlated), something that the previous literature has not been able to do. Overall, our results suggest that customer satisfaction acts as a reasonable proxy for service performance and provides additional informational value to firms as a pathway bridging the impact of objective service performance to customer purchase behavior. We use these results to show that failing to maintain good service performance leads to undesirable changes in customer purchasing behavior. Our results replicate across both service settings - quick service restaurants and auto rental - allowing to speculate that these results are generalizable. We demonstrate the behavioral impact of changes in service quality based on our results. Finally we carry out a set of robustness checks (using difference measures of dependent and independent variables, examining across-individual selection as well as explicitly controlling for product, as opposed to service, quality).

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 past literature. Next, we present our model and estimation procedure and describe the institutional setting, the data and the operationalization of the variables across the two different industries. We also 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 and Its Antecedents

Previous research defines 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 a transaction-specific reaction (Oliver 1980) or 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)) depending on the research context. The former references the customer's satisfaction with a specific, discrete service encounter and has typically been measured by asking survey participants to consider the last experience they had (e.g., Olsen and Johnson 2003). The latter refers to the customer's cumulative evaluation based on all encounters and experiences with the organization. Notably, the attitude-based satisfaction measure resembles the concept of perceived service quality in the service literature (Parasuraman et al. 1988). In this research, we focus on transaction-based satisfaction.

Early satisfaction research identifies disconfirmation as one of the key antecedents of customer satisfaction. Disconfirmation is the discrepancy between performance and expectation and has been viewed as a distinct and independent construct from (performance) expectation. Positive disconfirmation (when performance exceeds expectation) is hypothesized to increase customer satisfaction and negative disconfirmation to decrease customer satisfaction (Oliver 1980). Prior research has consistently found a significant effect of disconfirmation on customer satisfaction (in the direction noted above).[1] Some research also reports asymmetric disconfirmation effects where the negative disconfirmation effect is stronger than the positive counterpart (e.g., Anderson and Sullivan 1993).

Different approaches have been proposed to operationalize the disconfirmation construct. First, early research measured the objective discrepancy between expectations and performance outcomes in an experimental setup to derive a difference score (Weaver and Brickman 1974). Second, Parasuraman et al. (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 preserve 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). These self-reported measures of expectation 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,

[1] Previous research has found mixed or weak effects of expectation and perceived performance on customer satisfaction (e.g., Anderson and Sullivan 1993; Churchill and Surprenant 1982; Voss et al. 1998).

not survey or experimental, data.

Prior research has consistently found a significant effect of disconfirmation on customer satisfaction (in the direction noted above), but mixed or weak effects of expectation (e.g., Voss et al. 1998). Some research also reports asymmetric disconfirmation effects where the negative disconfirmation effect is stronger than the positive counterpart (e.g., Anderson and Sullivan 1993). At the same time, later research argues that customer satisfaction is more likely to be influenced by performance alone. For instance, Churchill and Surprenant (1982) show that for durable as opposed to non-durable products only perceived performance, and neither expectation nor disconfirmation, affects customer satisfaction. The explanation for this is that customers often do not have enough information to form a reliable expectation for an infrequently purchased durable product. Anderson and Sullivan (1993) also find that disconfirmation together with perceived performance affects customer satisfaction (instead of disconfirmation and expectation), based on a post-purchase survey covering a variety of product categories.

Given the availability of multi-period panel data of objective performance, satisfaction rating, and purchase behavior at the individual level, we specify the evolution of an individual's expectation to follow an anchoring and adjustment process (Boulding et al. 1993; Johnson et al. 1995; Kopalle and Lehmann 1995) and derive it as a function of the objective performance that varies over time. Note that the derivation of customer expectation based on objective service performance data helps us circumvent the mere-measurement effect. Previous research argues that prompting customers' expectations sensitizes them toward 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 in the control groups who had not. We then conceptualize that disconfirmation - the difference between the derived expectation and objective performance - ultimately affects transaction-based satisfaction.

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 (e.g., Godfrey et al. 2011). 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. As a result, previous customer satisfaction literature controls for the effect of unmeasured characteristics related to the selection process. For example, Godfrey et al. (2011) take into account *across-individual* selection bias by first modeling customer's propensity to be included in the satisfaction survey and then using the obtained inverse Mill ratios as a control variable that links customer satisfaction to repurchase. To examine the impact of firm's customer relationship management (CRM) activities on customer satisfaction, Mithas et al. (2005) use a propensity score matching approach to control for *across-individual* selection bias given rise from the researchers' inability to exogenously assign firms to the CRM treatment.

However, to the best of our knowledge, no previous customer satisfaction study has addressed the *within-individual* selection process in satisfaction rating. In our study, we obtain two unique datasets where we observe satisfaction ratings and purchases over multiple time periods. Individuals in our datasets had an opportunity to rate their satisfaction after each purchase transaction but they could also opt out. As such, an individual's satisfaction rating involves two decisions of whether to rate (i.e., rating incidence) and what rating to give (i.e., rating decision). Both of these decisions may be influenced by the individual's independent evaluation of the purchase experience and at the same time are interdependent. In particular, the unobserved factors driving customers' propensity to rate are likely to be correlated with the observed satisfaction rating. This situation gives rise to selection because the unobserved factors affect both rating incidence and actual ratings. Ignoring this *within-individual* selection bias will lead to incorrect inferences regarding the observable factors driving the actual ratings. We model the rating incidence and the actual satisfaction rating simultaneously to address the *within-individual* selection bias in customer satisfaction rating.

Objective versus Perceived Performance Measures

As objective measures of performance are hard or even infeasible to obtain (e.g., helpfulness of a salesperson, reliability) especially in the service context, most satisfaction research relies on perceived performance measures (e.g., Churchill and Surprenant 1982; Oliver 1980). Objective measures of performance can be based on either observable and concrete metrics (e.g., minutes, number of defects) or expert ratings (Mitra and Golder 2006). In contrast to objective performance, perceived performance is derived from customers' subjective judgment of the observed performance. As a result, perceived performance does not necessarily reflect actual performance as customer perception is likely to be influenced by factors such as marketing communication and experiences of others, as well as prior expectations of performance (Anderson and Sullivan 1993).

Several studies have managed to secure objective performance measures. Gijsenberg et al. (2015) find the proportion of successful connections in the railway service industry to affect aggregate customer satisfaction. Bolton et al. (2006) link the objective performance of a supplier's engineering service (e.g., work minutes per a support request), and Sriram et al. (2015) link signal quality of a video-on-demand service to customer retention. However, these papers do not study satisfaction in the same framework. In non-contractual service settings, Grewal et al. (2010) find a significant relationship between objective performance in the airline industry (e.g., percentage of on-time arrival, mishandled baggage and complaint) and overall attitude-based customer satisfaction. Lund and Marinova (2014) show that objective service performance in the pizza restaurant industry (i.e., delivery time) negatively moderates the impact of direct marketing effort on retail revenue. In the last two papers, however, objective service performance measures are not obtained at the transaction level.

To the best of our knowledge, our research is the first to conduct an individual-level analysis in a transaction-based (non-contractual) setting to examine the interplay among objective performance, transaction-based satisfaction and purchase behavior in the same framework. Prior research largely relies on self-report ratings of satisfaction, perceived performance, expectation, disconfirmation and purchase intent to investigate the antecedents and outcome of satisfaction. It is widely recognized that such approach is prone to high common-methods variance, which likely inflates the relationships among the constructs under investigation (Fishbein and Ajzen 1975). Our research matches satisfaction ratings to observed objective performance and purchase transactions over time and hence avoids such a problem.

Impact of Satisfaction on Purchase Behavior

Satisfaction research has consistently shown the impact of satisfaction on purchase intention (Anderson and Sullivan 1993; Oliver 1980) and on downstream business outcomes such as service usage (Bolton 1998), retention (Rust and Zahorik 1993), share of customer wallet (Bowman and Narayandas 2001), and firm's financial performance (Luo et al. 2010). Nonetheless, some research finds the direct main effect of satisfaction on individual-level purchase behavior (Seiders et al. 2005) to be insignificant. The absence of the direct effect can be attributed to the subtle relationship between satisfaction and purchase behavior. First of all, some studies argue that the relationship between satisfaction and repurchase is nonlinear and asymmetric. For instance, Mittal and Kamakura (2001) find a convex relationship between customer satisfaction and repeat purchase, with customer satisfaction changes at the top end of the scale having the biggest impact. Second, the impact of satisfaction on purchase behavior can be moderated by customer characteristics (e.g., usage level, firm size), the strength/age of customer relationship with the firm, and marketplace characteristics such as competition and satisfaction with competitors (Seiders et al. 2005).

To establish the link between customer satisfaction and purchase behavior, some prior research administered a survey to measure customer satisfaction at T_0 and later observed the customer's purchase behavior at T_1 as they have no access to the customer database (Rust and Zahorik 1993) or wish to also obtain information about the customers' purchases of competitors' products (Verhoef 2003). All this research encounters the usual problem of declining survey response rates as the number of surveys increases. Unlike most previous research that relies on cross-sectional survey data, we observe satisfaction ratings and purchases in multiple periods for each individual. Our behavioral variable of interest is, similar to Bolton (1998), the actual interpurchase time at the individual level.

Performance Consistency

The service management literature has emphasized the importance of consistent service performance in maintaining high perceived quality, service value and customer loyalty. Reliability - partly described as consistency in performance - is found to be an important determinant of overall perceived service performance (Parasuraman et al. 1988). For example, McCollough et al. (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. In their longitudinal experiment, 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 over time.

Other research provides some caveats that some particular situations of performance inconsistency may also lead to positive outcomes. Bolton et al. (2006) find a few extremely favorable experiences to be critical for business customers' subsequent re-patronage behavior such as system support contract renewal. Similarly, Hansen and Danaher (1999) find the end experiences within a service encounter to increase satisfaction. Sriram et al. (2015) show that high levels of service variability can increase customer retention when the general service performance is low.

To measure performance inconsistency, the previous literature in service performance has used different measures e.g., the number of extremely positive or negative performance (Bolton et al. 2006), the proportion of successful performance (Gijsenberg et al. 2015), and the variance of service performance (Sriram et al. 2015). Taking advantage of multi-period panel data of objective performance, this paper specifies the evolution of performance inconsistency by calculating the cumulative standard deviation of objective service performance up to each time period. We expect to find performance inconsistency to negatively affect customer satisfaction and actual purchase behavior.

MODELING APPROACH

Model Specification

We model satisfaction rating incidence, satisfaction rating, and interpurchase time as three separable but related processes, by constructing a set of simultaneous equations at the individual level. In the first step, we consider two decisions by the individual at each service encounter: whether to provide satisfaction ratings and if so, what ratings to give. We use a binary Probit and an ordered Probit to model rating incidence and satisfaction rating, along with a correlated error structure between the two models. Given that observed satisfaction ratings are conditional on rating incidence, ignoring rating incidence or treating satisfaction rating as being independent of rating incidence can give rise to a selection bias. To circumvent this *within-individual* selection bias, we need to account for the potential unobserved factors (e.g., competitors' promotional activities) that affect both rating incidence and satisfaction rating (see Narayanan and Manchanda 2012 for a similar situation in a different institutional setting). In the second step, we investigate the impact of customer satisfaction on interpurchase time. We regress log-transformed interpurchase time on the *predicted satisfaction rating generated from the first step* as that represents the "unbiased" rating, as well as other variables of interest.[2] For customer i's purchasing at store j on purchase occasion t, the system of equations is specified as follows:

$$(1) \quad INC^*_{ijt} = \alpha' Z^{INC} + \epsilon^{INC}_{ijt}, \quad INC_{ijt} = 1 \text{ where } INC^*_{ijt} > 0, \; INC_{ijt} = 0 \text{ otherwise}$$

$$(2) \quad SAT^*_{ijt} = \beta' Z^{SAT} + \epsilon^{SAT}_{ijt}, \quad SAT_{ijt} = \kappa \text{ where } \mu_{\kappa-1} < SAT^*_{ijt} \leq \mu_\kappa, \; \kappa = 1, \cdots, 5$$

$$(3) \quad INT_{ijt,t+1} = \gamma_1 X + \gamma_2 \widehat{SAT}_{ijt} + \epsilon^{INT}_{ijt}$$

where INC^*_{ijt} and SAT^*_{ijt} are the underlying latent variables representing the customer's decision of whether to rate and what rating to give if she provides the rating. INC_{ijt}, SAT_{ijt}, and $INT_{ijt,t+1}$ are observed rating incidence, satisfaction ratings, and log-transformed interpurchase time in the data, respectively. $\widehat{SAT}_{ijt}$ is the predicted customer satisfaction generated from Equation 1 and 2. $Z^{INC} = \{DIS_{ijt}, VAR_{ijt}, CPN_{ijt}, NTR_{ijt}\}$, $Z^{SAT} = \{DIS_{ijt}, VAR_{ijt}, CPN_{ijt}\}$, and $X = \{DIS_{ijt}, VAR_{ijt}, CPN_{ijt}, INT_{ijt,t-1}, AMT_{ijt}\}$ represent sets of explanatory variables for rating incidence, satisfaction rating, and interpurchase time, respectively. Following the previous studies (Weaver and Brickman 1974) that take a more objective approach, we operationalize disconfirmation, DIS_{ijt}, as the difference between current objective service performance and the customer's prior expectation on the performance.[3] Customer expectations are updated based on objective service performance and prior expectations in the previous periods (e.g., Kopalle and Lehmann 1995; Johnson et al. 1995). In this adaptive expectation setup, the greater the weight on the objective performance, the more significant the effect of immediate past experience on current

[2] Alternatively, the probability of purchase conditional on interpurchase time could be modeled as a hazard function. We chose a linear model of log-transformed interpurchase time because of its simplicity.

[3] Other research has used survey based measures of expectation. For example, Boulding et al. (1993) propose 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 will expectation increases perceived (as opposed to objective) quality, while should expectation does the opposite. However, they do not discuss the link between these two types of expectation and customer satisfaction. The use of survey based expectation, objective quality and the lack of the link between expectation and customer satisfaction makes it hard for us to investigate the role of these two forms of expectation in our data in an "apples-to-apples" comparison. For example, preliminary analyses with a measure for should expectation in our setting provides results that are not consistent with those in Boulding et al. (1993). We speculate that this is due to the differences in the expectation measures - our measures are objective and transaction specific while the measures in Boulding et al. (1993) are subjective, based broadly on the customers' *overall* experience with the firm (and its competitors).

expectation, or the more adaptive the expectations (Johnson et al. 1995).

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

where $PERF_{ijt}$ is objective service performance and EXP_{ijt} is the customer's expectation on the performance. 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 smoothing factor δ, we perform a grid search (over the [0, 1] interval with a step size of 0.1).

Service performance inconsistency, VAR_{ijt}, is operationalized as a cumulative standard deviation of delivery time up to the current service encounter (e.g., Sriram et al. 2015). Note that the first two observations for each individual are not included in the model estimation but used to calculate the cumulative standard deviation of objective service performance. In addition, we include customer coupon redemption, CPN_{ijt}, and dollar purchase amount, AMT_{ijt}, as control variables. To address the endogenous relationship between rating incidence and satisfaction rating, we use the number of transactions since the last time the customer provided a satisfaction rating, NTR_{ijt}, as an exclusion restriction. We assume that (either negative or positive) service performance spark customers' interest and initiate their participation in a survey (Dellarocas and Narayan 2006). When being repetitively asked to take the same survey, however, customers are likely to become satiated (Bickart and Schmittlein 1999). Previous research finds that over-surveying results in lowered response rates because (1) with increasing contacts, respondents' overall attitudes toward the survey may become less favorable, and (2) as people are contacted more often, they feel that the opportunity to provide their opinions in a survey is not a "rare" and, therefore, no longer a valuable experience (Groves et al. 1992).

The cut-points for the ordered Probit in Equation 2, $\{\mu_0, \cdots, \mu_5\}$, determine how the latent scale, SAT^*_{ijt}, is mapped onto the observed ordinal scale points. For identification purposes, we set the lowest and the highest cutoffs (μ_0 and μ_5) to $-\infty$ and ∞, and the second lowest cutoff (μ_1) is fixed to zero (e.g., Ying et al. 2006). Note that we cannot capture customer heterogeneity in the cut-points directly because they are positive and obey order restrictions. Instead, we allow the latent utility (Equation 2) to contain a random intercept and estimate the differences between the cut-points and the intercept. The error terms in Equation 1 and 2, ϵ^{INC}_{ijt} and ϵ^{SAT}_{ijt}, are assumed to have a multivariate normal distribution with mean vector of zero and covariance matrix of $(1, 1; \rho_{12})$. The error structure between the rating incidence and the satisfaction rating equations explains the correlation between unobserved components in customer rating behavior and controls for the *within-individual* selection problem. We also 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 the customer provided a satisfaction rating as an exclusion restriction. We assume that (either negative or positive) service performance spark customers' interest and initiate their participation in a survey (Dellarocas and Narayan 2006). When being repetitively asked to take the same survey, however, customers are likely to become satiated (Bickart and Schmittlein 1999). Previous research finds that over-surveying results in lowered response rates because (1) with increasing contacts, respondents' overall attitudes toward the survey may become less favorable, and (2) as people are contacted more often, they feel that the opportunity to provide their opinions in a survey is not a "rare" and, therefore, no longer a valuable experience (Groves et al. 1992).

We use random coefficients to control for unobserved heterogeneity at the individual level. Specifically, we allow the intercept terms and the coefficients of disconfirmation and performance inconsistency to vary across household. This specification also helps us take into account unobserved customer heterogeneity for rating incidence and rating decision. We also incorporate a similar random coefficient specification into the interpurchase time model where we allow the intercept and the coefficients of the predicted customer satisfaction, disconfirmation, and performance inconsistency to be different for each individual. 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 because the orders from each customer are almost always confined to a certain store based on his or her address.

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) to address the within-individual selection bias. We estimate the two equations simultaneously, using a partial maximum likelihood approach (Popuri and Bhat 2003):

(5)

$$\log L = \sum_{INC=0} \log \Phi(-\alpha' Z^{INC}) + \sum_{INC=1} \sum_{k=1}^{5} m_{ijtk} \log \left[\Phi_2(\mu_k - \beta' Z^{SAT}, \alpha' Z^{INC}, \rho_{12}) - \Phi_2(\mu_{k-1} - \beta' Z^{SAT}, \alpha' Z^{INC}, \rho_{12})\right]$$

where $m_{ijtk} = 1$ if $SAT_{ijt} = k$ and Φ_2 represents a cumulative bivariate normal density. The correlation between rating incidence and satisfaction rating operates through the nonzero correlation, ρ_{12}. Because ρ_{12} is bounded between $[-1, 1]$, we use the Fisher transformation, $arctanh(\rho_{12}) = (1/2)\ln[(1+\rho_{12})/(1-\rho_{12})]$, and map $[-1, 1]$ to the real line (Ying et al. 2006). where $m_{ijtk} = 1$ if $SAT_{ijt} = k$ and Φ_2 represents a cumulative bivariate normal density. The bivariate normal distribution of the errors terms helps capture the potential correlation in unobserved factors that simultaneously drive both rating incidence and actual satisfaction rating (Greene 2011). The correlation between rating incidence and satisfaction rating operates through the nonzero correlation, $\rho_{12} = [-1, 1]$. We use the Fisher transformation, $arctanh(\rho_{12}) = (1/2)\ln[(1 + \rho_{12})/(1 - \rho_{12})]$, and map $[-1, 1]$ to the real line (Ying et al. 2006). In our context, constructing Inverse Mills ratio from the selection equation and adding it to the outcome equation (Heckman 1979) might result in biased estimates because our satisfaction equation (Equation 2) is nonlinear. Note that the Heckman approach assumes that the distribution of ϵ_{ijt}^{INC} is known and ϵ_{ijt}^{SAT} is a linear function of ϵ_{ijt}^{INC}.

We use the simulated maximum likelihood approach (Greene 2011) to obtain the random effects. Specifically, we take R draws from a standard normal density, u_{ir} where $r = 1, \cdots, R$, and create the sampling distribution of the parameters to be estimated. The resulting values are $\alpha^r = \hat{\alpha} + \sigma_{INC} u_{ir}$ and $\beta^r = \hat{\beta} + \sigma_{SAT} u_{ir}$, where $\hat{\alpha}$ and $\hat{\beta}$ are the population parameters. σ_{INC} and σ_{SAT} represent the unobserved, individual specific heterogeneity. The following simulated log-likelihood function is obtained by integrating out the unobserved variable u_{ir}:

$$\begin{aligned}
\log L_{SML} = & \sum_{i=1}^{n} \log \Big\{ \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} \Big[\prod_{t=1}^{T_i} \Phi(-\alpha^{r\prime} Z^{INC}) \Big] \Big\} \\
& + \sum_{i=1}^{n} \log \Big\{ \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} \Big[\prod_{t=1}^{T_i} \prod_{k=1}^{5} m_{ijtk} \big[\Phi_2(\mu_k - \beta^{r\prime} Z^{SAT}, \alpha^{r\prime} Z^{INC}, \rho_{12}) \\
& \qquad - \Phi_2(\mu_{k-1} - \beta^{r\prime} Z^{SAT}, \alpha^{r\prime} Z^{INC}, \rho_{12}) \big] \Big] \Big\}
\end{aligned} \tag{6}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ is the standard normal PDF and CDF. The parameters to be estimated are $\{\hat{\alpha}, \hat{\beta}, \sigma_{INC}, \sigma_{SAT}\}$. We set $R = 100$.

With the parameter estimates from Equation 6 in hand, we obtain a consistent asymptotically normal estimator of customer satisfaction and use the value to estimate the impact of customer satisfaction on interpurchase time (i.e., Equation 3). Similar to Terza (1987), we develop the following predicted customer satisfaction measure as a regressor in the interpurchase time equation:

$$\widehat{SAT_{ijt}} = \sum_{k=1}^{5} m_{ijt} \Big[\frac{\phi_2(\mu_{k-1} - \beta' Z^{SAT}, \alpha' Z^{INC}, \rho_{12}) - \phi_2(\mu_k - \beta' Z^{SAT}, \alpha' Z^{INC}, \rho_{12})}{\Phi_2(\mu_k - \beta' Z^{SAT}, \alpha' Z^{INC}, \rho_{12}) - \Phi_2(\mu_{k-1} - \beta' Z^{SAT}, \alpha' Z^{INC}, \rho_{12})} \Big] \tag{7}$$

where $m_{ijtk} = 1$ if $SAT_{ijt} = k$. $\phi(\cdot)$ and $\Phi(\cdot)$ indicates the standard normal PDF and CDF. This functional form provides consistent results with the fitted value from the satisfaction equations (Equation 1 and 2) while circumventing the multicollinearity problem in the interpurchase time equation (Equation 3).

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 aspect, especially delivery time, of the transaction influences both customer satisfaction and purchase behavior (Verma et al. 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 (e.g., 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: 1) when the order is placed, 2) when the order comes out of the oven, 3) when the driver leaves the store, and 4) when the driver returns to the store. The delivery time for each order is calculated to be [((the 4th time stamp - the 3rd time stamp)/2) - the 1st time stamp] + 2 minutes.[4] 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 determinant of customer satisfaction and business outcomes. However, in this industry (and in our setting), none of these are obtained at the transaction level (e.g., Lund and Marinova 2014).

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 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 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). Thus, the focus in this study is transaction-based, rather than attitude-based, satisfaction. 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?

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

- 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

Transaction data. The data span a total of 1,505,529 delivery orders from 362,672 unique customers (households)[5] who provided satisfaction ratings at least once during the sample period at 55 stores in Texas and Virginia from January to December 2011.[6] 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 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 over time (27.5% of carryout-only and 62.1% of delivery-only customers). 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 6.85% of all the transactions where customers order from different stores over time. As we derive expectation, disconfirmation and performance consistency 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 consistency and link it to purchase behavior in T_2. We thus limit the sample to 1,257,174 transactions from 177,922 customers who purchased three or more times. As can be seen from Table 1, there is no significant difference in behavior between the households in sample with at least three purchases and the entire sample.

[Insert Table 1 about here]

Satisfaction Ratings. From the six questions on the online survey, we use Q1 ("How likely are you

[5]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 "individual" and "household" interchangeably. This remains a limitation of our approach.

[6]The overall proportion of delivery orders is 59.8%. We also have an additional 7,014,948 delivery orders from 5,042,059 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-individuals* selection bias in the "Robustness Checks" section.

to recommend us to your family and friends?") as the measure of transaction-specific customer satisfaction. Previous research finds recommendation and satisfaction items to form an overall assessment of customer satisfaction. For example, Keiningham et al. (2007) show that customer responses to a willing-to-recommend survey question are very similar to satisfaction measures (the American Customer Satisfaction Index). Additionally, Q2 and Q3 most likely reflect customer evaluation on ordering methods (e.g., how cumbersome and time consuming to place an order via phone or online). Q4 and Q5 also appear to capture idiosyncratic responses about the driver or clerk, not the delivery time. Q6 measures customers' perceived delivery time, instead of customer satisfaction.[7] Finally, factor analysis using orthogonal Varimax rotation shows that responses to Q1 load on a different factor from responses to Q3, Q4, and Q6. Figure 1 shows the distribution of Q1 and the average delivery time for each rating. As shown, the ratings are skewed towards the 4 and 5 scores and our focal objective service performance - delivery time - is negatively correlated with satisfaction ratings (the correlation coefficient is -0.11 and a regression of satisfaction ratings on delivery times shows that the latter has a significant and negative coefficient).

[Insert Figure 1 about here]

Customer satisfaction ratings were provided in 2.8% of all transactions. While this may seem low, this is consistent with industry numbers (based on feedback from the data provider). Conditional on individuals providing a rating (during the sample period), participation rates went up quite dramatically with 16.0% of transactions being rated, i.e., 1.73 times per individual on average.

Sample and Summary Statistics. In our empirical analysis, we randomly sample 20,000 (out of 177,922) customers with 141,301 transactions, who purchased three or more times and provided satisfaction ratings at least once during the 12-month period. Table 2 reports summary statistics on delivery time, coupon redemption, purchase amount, and interpurchase time. Table 2 presents the summary statistics (1) from the transactions with satisfaction ratings and (2) from transactions without satisfaction ratings.

[Insert Table 2 about here]

As can be seen from Table 2, there is no significant difference in disconfirmation and performance

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

inconsistency across the three different samples. In contrast, interpurchase time is longer and the number of transactions since the last rating is smaller for transactions with satisfaction ratings relative to those without ratings, while the differences are not statistically significant. 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.

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 3 presents the key variables of the transactions made by (1) "raters" and (2) "non-raters." The descriptive statistics suggest that the differences between "raters" and "non-raters" on the key metrics (e.g., interpurchase time of 36.8 vs. 41.0 days) 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).

[Insert Table 3 about here]

Results

Customer Rating Behavior. In this section, we show the results from our proposed model of customer rating behavior which includes both the decision to rate and the actual rating, conditional on the rating decision. Table 4 reports the results from two different specifications: (1) a null model where we ignore within-individual selection (i.e., only use an ordered Probit model for satisfaction rating) and (2) a simultaneous equation model of rating incidence and satisfaction rating to control for within-individual selection with unobserved customer heterogeneity using the random effects specification. Model 2 is our proposed model.

[Insert Table 4 about here]

Overall, we find that objective service performance does have a clear impact on customer satisfaction rating, thus answering our first research question. The parameter estimates in Table 4 suggest that both disconfirmation and performance inconsistency are key determinants of customer satisfaction rating. First, higher disconfirmation (delivery time is longer than expected) decreases customer satisfaction rating. These results confirm findings from the previous literature that proposes a relationship between disconfirmation and customer satisfaction (e.g., Oliver 1980).[8] Model 2 confirms that there is a selection bias in within-household ratings over time as the error correlation, ρ_{12}, is significant and positive, and as such, it is important to correct for the within-individual selection. The positive correlation suggests that customers are more likely to provide rating when they feel more positive about the service they received. The number of transactions since the previous rating - the proposed instrument to identify the selection process - significantly decreases customers' participation in satisfaction ratings. The more recently customers have rated, the less likely they are to provide ratings again. To the best of our knowledge, results linking objective service performance and satisfaction ratings while accounting for *within-individual* selection have not been presented before.

Customer Repurchase Behavior. Next we focus on our second research question of whether customer satisfaction provides additional information over and above the information present in objective service performance. We estimate our proposed model where we link predicted customer satisfaction (as these ratings are corrected for selection) along with disconfirmation and performance inconsistency to interpurchase time using only those observations for which we have satisfaction ratings. We then compare our proposed model with a series of alternative models in order to answer our research question. Table 5 reports the results.

[Insert Table 5 about here]

The results based on the proposed model (Model 1) show that objective service performance does not directly impact interpurchase time. Customer satisfaction, however, has a direct impact on interpurchase time with higher satisfaction leading to shorter interpurchase time. Another way to think about this is that, over and above the direct impact, disconfirmation and performance inconsistency indirectly affect interpurchase time through customer satisfaction. In other words, the impact of objective service performance on customers' interpurchase time is operated through the satisfaction rating. An alternative explanation is that the direct impact of customer satisfaction on repurchase might be attributed to the mere-measurement effect where measurement of customer

[8] We also find an asymmetric effect of disconfirmation on customer satisfaction in a separate model. Following Anderson and Sullivan (1993), we test for the asymmetric effect by including terms to represent differential effects of negative disconfirmation (delivery time is shorter than expected) versus positive disconfirmation (delivery time is longer than expected) on satisfaction. The results show that satisfaction is much more sensitive to positive than negative disconfirmation ($\beta_{positive} = -0.9632$ vs. $\beta_{negative} = 0.0061$). Detailed results are available from the authors on request.

intentions or customer participation in surveys could positively influence customer retention (Dong et al. 2014). This explanation is unlikely in our context because the correlation between survey participation and interpurchase time is very small (0.04) and in an opposite direction as suggested by mere-measurement literature.

Our proposed model uses constructed and/or transformed measure such as predicted satisfaction, 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 customers having provided satisfaction ratings, the satisfaction metric (the raw subjective measure of service performance) does have the expected relationship with interpurchase time - higher satisfaction leads to quicker purchases. The results from Model 3, on the other hand, suggest that there is a weak direct relationship (the correlation coefficient is 0.02) between interpurchase time and delivery time (the raw objective measure of service performance).

Models 4 and 5 also use constructed measures but allow us to examine different ways of using the objective and subjective performance data. Model 4 uses only disconfirmation and performance inconsistency (without the predicted satisfaction) measures; the results are qualitatively similar to those from Model 1. Finally, Model 5 uses disconfirmation and performance inconsistency together with the residual obtained from the proposed satisfaction rating model (Model 2 in 4). This residual represents the information contained in the satisfaction rating over and above that in the observables in our model. Disconfirmation, performance inconsistency, and the residual term are not significant. In other words, despite its role in bridging objective service performance and customer repurchase, customer satisfaction does not provide information over and above what can be explained by the observables.

Turning to model fit, we find that our proposed model provides the best in-sample fit. We also examine the models' predictive performance using a holdout sample. Specifically, we re-estimated the models without the last transactions of each customer and use them as our holdout sample. Our fit statistics are root mean square error (RMSE), mean standard deviation (MAD), and mean absolute error (MAE). Our proposed model has the best predictive performance with respect to all three fit statistics.

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 customers' repurchase behavior through customer satisfaction, and therefore provides implementable suggestion of how firms can improve their service. In the "The Effect of Delay in Service" section, we further explore the impact of managerial actions on firm performance.

Robustness Checks

In this section, we report results from a series of robustness checks. First, we investigate the impact of satisfaction rating and objective service performance on purchase amount instead of interpurchase time. Second, we explore the impact of objective service performance on customer satisfaction and interpurchase time controlling for product performance. Third, we examine the relative magnitude of across-individuals and within-individual selection biases. Finally, we test alternative measures of customer expectation to calculate disconfirmation.

Customer Purchase Amount. We test to see if our results are robust to an alternative measure of business outcome, in our case, the dollar amount of each order. Similar to the approach in Equation 3, we treat the predicted measure of customer satisfaction as an independent variable in the regression with the dependent variable being the log-transformed purchase amount. The results are reported in Table 6 (Column (1)). Similar to the interpurchase time model in Table 5, high predicted satisfaction rating leads to significantly higher purchase amount, while the direct impact of objective service performance is not observed. These results confirm the role of satisfaction rating in bridging objective service performance and customer purchase behavior using an alternative outcome measure, providing convergent validity and providing us confidence in our findings.

[Insert Table 6 about here]

Product Image Evaluations. In the analysis so far, we have assumed that our measure of service performance is invariant to the quality of the delivered product. However, we also have a unique opportunity to look at the effect of service performance while explicitly controlling for product performance. This is because, in selected stores, the company has installed a camera to take pictures of food coming out of the oven.[9] Thus, in addition to the satisfaction measure and the tracking of delivery time, the company collects product image evaluations from these stores yielding (novel) measures of product performance.

[9]The proportion of franchise stores in the sample of the stores with the camera - about 62% - is consistent with that in the entire sample. Note that no changes in the recipe or the formulation occurred during the span of our data.

For each of these stores, each order is photographed each day as it comes out of the oven. Five product images are then randomly picked from each store to be rated by a team of hired raters on a binary scale of "good" or "bad." The raters are not aware of which store's products they are scoring. The company has analyzed the scores to provide feedback to franchisees with regard to their food quality and has found that the scores are a valid predictor of store performance e.g., stores with relatively low product performance scores are likely to have sales that were lower than expected, and drops or increases in product performance scores tend to correlate with (delayed) sales shifts. Figure 2 shows the distribution of average image scores (where 1 is "good" and 0 is "bad") and the distribution of interpurchase time across the different average scores. The product image evaluations during the sample period are reasonably high and they do not appear to be correlated with interpurchase time (correlation coefficient -0.03). Note that the average image ratings per day are clustered on 0.2, 0.4, 0.6, 0.8, and 1.0 as each store gets five evaluations a day.

[Insert Figure 2 about here]

To control for product performance in customer satisfaction and repurchase, we re-run our proposed model using only the service transactions where product image evaluations are available. In this analysis the store-level daily average score (based on five randomly selected image scores) is used as the measure of product performance for each store on a given day. The small sample of stores with product image evaluations leads to a much smaller number of observations for this analysis. Despite the small sample, as shown in Table 6 (Column 2), the results again demonstrate the value of satisfaction in linking objective service performance to purchase behavior. We also observe within-individual selection. With this much smaller sub-sample, however, we observe the direct impact of disconfirmation and performance inconsistency on interpurchase time, which differs from the findings from our proposed model. The discrepancy between the results in this analysis and those in Table 4 and 5 is thus also likely driven by the difference in sample sizes. Note that we do not find product performance to significantly affect either satisfaction or interpurchase time.

Selection Problem across Individuals. In our analysis so far, we have accounted 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 a different 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 our sample. To match the same number of observations we use in the models reported in Table 4 and 5, we draw a sample of 141,301 customers from the population of all customers who made at least three purchases during the sample 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 industry and the previous literature to obtain satisfaction ratings. Table 6 (Column (3)) reports the results of the model estimated on this sample.

The results show that across-individual selection is not significant and, at the same time, its magnitude is not as large as that of within-individual selection. These results suggest two things - first, the behavior of "raters" and "non-raters" are not significantly different and second, selection remains an important issue vis-á-vis satisfaction ratings within household.

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 alternative measures of customer expectation. Customers may imperfectly recall their prior service performance because of factors such as high cognitive efforts required for adjusting prior expectation, low involvement, and low purchase frequencies (e.g., Mitra and Golder 2006). Consistent with the literature, the exponential smoothing approach in our proposed model gives higher weights to service performance that occurs more recently. We examine two other specifications that assume either a shorter-term memory or a perfect memory. For the former, we consider the immediate past service encounter as a measure of customer expectation. For the latter, we use a simple moving average where all past service performance contributes equally to customers' expectation. We then calculate disconfirmation using Equation 4. The results in Table 6 report that the effect of disconfirmation is robust across different measures of customer expectation. As shown in Column (4) and (5), no direct impact of disconfirmation is observed with respect to rating incidence and inter-purchase time. Positive disconfirmation (i.e., worse-than-expected service performance) only decreases satisfaction rating.

The Effect of Delay in Service

One of the important issues that our research attempts to investigate is the impact of objective service performance on customer repurchase. For instance, we have already seen that disconfirmation and performance inconsistency significantly affect (lengthen) interpurchase time indirectly through customer satisfaction. Thus, minimizing customers' disconfirmation and performance inconsistency would be crucial for firms to shorten the customer's interpurchase time. A delay in service that increases both disconfirmation and performance thus worsens economic outcomes for the firm.

We explore this effect of delay in service via a simulation where we manipulate delivery time, our key objective service performance. In particular, we specify "always worse service" where delivery time increases by k% each period (with maximum of 2 hours), setting $TIME_t = (1 + k\%) \times TIME_{t-1}$, where $k = 5, 10, \cdots, 50$. Next, we update disconfirmation and performance inconsistency based on the simulated delivery time. Finally, we run our proposed model and investigate the impact of improvement in delivery time on interpurchase time and purchase amount. Figure 3a show that "always worse service" hurts the firm. In particular, customers purchase less frequently as delivery time increases. For example, if service performance worsens by 20% from the previous service encounter, the firm could expect customers' interpurchase time to increase by approximately 1.6%. The decreasing effect of service performance on interpurchase time indicates that the disconfirmation effect overrides the effect of performance inconsistency on interpurchase time, as the firm continues to decrease service performance. We repeat the simulation to investigate the impact of decrease in objective service performance on customer satisfaction ratings. The proportion of "extremely" or "very" satisfied customers (i.e., satisfaction rating = 5) substantially decreases from more than 90% to approximately 77% (at a 50% increase in delivery time), while the distribution of satisfaction ratings is still skewed towards the highest score (see Figure 3b).

[Insert Figure 3 about here]

STUDY 2: AUTO RENTAL INDUSTRY

Institutional Background

In our second study we use the data from a major car rental company in the United States. 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, in which customers receive a higher car-class for no extra charge. When demand for a lower car-class exceeds the available inventory and the forecasted demand for higher-car class is low, auto rental companies provide free upgrades, using unutilized higher-class cars as a "cheap" way to avert customer complaints. The previous service literature identifies free upgrades as one of the (compensatory) recovery strategies to address service failure (e.g., Hoffman et al. 2003), which potentially increase customers' positive reactions. Other studies (e.g., Jiang et al. 2013) also show that providing customers with unearned preferential treatments such as a surprise discount and a free upgrade can generate increase customer satisfaction. Therefore, we use free car-class upgrade as the key objective service performance measure in our second analysis.

The auto rental firm tracks transaction-specific customer satisfaction from an online survey. In order to complete the survey, customers were provided with a hyperlink in one of the two ways: through email or on their printed receipt. Because the customers make satisfaction assessments after their rental experiences, we assume that the customer satisfaction ratings collected reflect transaction-based satisfaction, which is affected by the most recent service performance the customers have encountered. The 10-point scale (except Q3: 5-point scale) satisfaction survey consists of the following eight questions:

- 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

Transaction Data. The data used in this study cover the 4,386,650 of rental car transactions from

540,040 unique loyalty club members from October 2010 to September 2012 in the US.[10] Across 689 airport and 2,759 non-airport locations, 1 to 68 different car-classes are offered and the five most popular car-classes (i.e., Economy (A), Compact (B), Intermediate (C), Standard (D), Fullsize (F)) account for 90.1% of the total rental transactions. For each transaction, the firm collected data on the car class each customer reserved, the class the customer actually drove, and the class the customer paid for. When the reserved and the paid classes are same but the driven class is an upper-class vehicle, the customer receives a better class of car with no extra charge potentially because the reserved car is not available. This is called a car-class upgrade. Upgrades were offered 55.3% of the total transactions in the data. We also have the transaction detailed information such as store ID, customer ID, check-in/check-out date, base rental price, and satisfaction ratings. 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). We focus on customers who rented three or more times during 2-year sample period. As can be seen from Table 7, we do not observe substantial differences in the behavior between the sample of customers who purchased three or more times and the complete sample.

[Insert Table 7 about here]

Satisfaction Ratings. From the eight questions in the survey, consistent with the measure used in the previous study, we use Q2 ("How likely is it that you would recommend Hertz to a friend or colleague?") as the measure of transaction-specific customer satisfaction. Q2 is likely to reflect customers' evaluation on overall service performance including free car-class upgrades, while the other questions likely measure customers' idiosyncratic opinions about the staff at the counter, product performance, and transaction details. Factor analysis using orthogonal Varimax rotation suggests responses to Q2 load on a different factor from the responses to Q4, Q5, Q6, Q7, and Q8. While we acknowledge that Q1 and Q3 are possibly measures of overall service performance, we do not use customer responses to both questions due to the following issues with the data. First, we observe a very high proportion of raters (37.5%) gave the company the lowest rating on Q1,

[10] 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 ($>$ 99th percentile) of rental duration, reservation date, daily rental price, and purchase frequency.

something that we find implausible and inconsistent with all the other measures. We suspect that there could be some kind of measurement error. Second, over 60% of the responses to Q3 are missing mostly because the question was phased out in the middle of the data sampling period. Figure 4 shows the distribution of Q2 ratings and free upgrades for each rating. The satisfaction ratings during the sample period are reasonably high. Notably, more free upgrades are offered to the service encounters with higher customer satisfaction ratings (the correlation coefficient is 0.06 and a regression of satisfaction ratings on free upgrades shows a significant and positive coefficient). Customers provided their satisfaction ratings 1.38 times on average, which are 8.0% of the total transactions.

[Insert Figure 4 about here]

Sample and Summary Statistics. Consistent with Study 1, we randomly sample 20,000 (out of 304,516) customers who purchased three or more times. As a result, the sample size used in the model estimation is reduced to 267,713 transactions. Table 8 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 the two samples.

[Insert Table 8 about here]

From Table 8, we observe a very similar data pattern as did with Study 1. For example, average interpurchase times are substantially different across the two samples, where the transactions with satisfaction rating have the longer interpurchase time and the smaller number of transactions since the last rating, compared to those without ratings. The differences between the two samples open up a possibility of the within-individual selection problem. We thus use the transactions both with and without satisfaction rating to correct for selection. Note that customers in this study are all "raters," who participated in the survey at least once during the 24-month sample period. We therefore cannot examine across-individual selection.

Results

To explore customers' rating behavior in the auto rental industry, we take the same approach as that in Study 1, where we first estimate customer satisfaction rating behavior which includes the decision to rate and conditional on that decision, what ratings to give. We then answer our second research question of whether customer satisfaction ratings still provide information on repurchase behavior even in the presence of objective service performance data. Table 9 shows the results for (1) a null model where within-individual selection bias is not corrected and (2) our proposed model where the selection bias is addressed by a simultaneous equation of rating incidence and satisfaction rating. In both specifications, unobserved customer heterogeneity is controlled by the use of random coefficients.

[Insert Table 9 about here]

As shown in Table 9, our finding is consistent with Study 1 with respect to within-individual selection. The results from our proposed model (Model 2) indicates that within-individual selection needs to be addressed because the error correlation is significant. That is, more satisfied customers are more likely to rate. Different from Study 1, we also find a significant impact of objective service performance on the rating incidence. In particular, "lower than expected" service performance (i.e., negative disconfirmation) leads customers to participate in satisfaction survey. Regarding the impact of objective service performance on satisfaction ratings, consistent with Study 1, positive disconfirmation (i.e., offered an unexpected upgrade) significantly increases and performance inconsistency marginally decreases customer satisfaction (see Tables 10 for the summary). Our instrument to identify the selection process, the number of transactions since the previous rating, significantly decreases customers' participation in satisfaction rating, which is also consistent with the results from Study 1.

The parameter estimates in the interpurchase time equation confirm our findings from Study 1 (see Tables 10 for the summary). In particular, neither disconfirmation nor performance inconsistency significantly lengthens interpurchase time. Customer satisfaction, however, has a direct impact on interpurchase time with higher satisfaction leading to quicker purchases. In other words, we observe the role of customer satisfaction as a bridge between objective service performance and customer repurchase.

[Insert Table 10 about here]

The Effect of Car-Class Upgrades

As we did in Study 1, we examine the impact of improving customer service on interpurchase time. To do this, we manipulate the frequency of car-class upgrades (the objective service quality measure of interest). First, we randomly draw the car-class upgrade variable from a binomial distribution such that the probability of a car-class upgrade increases from the current level of 55% to 100%. We then update disconfirmation and performance inconsistency based on the simulated car-class upgrades and predict the resulting change in interpurchase time (see Figure 5). The results suggest that an increase in the proportion of car-class upgrades relative to current levels leads to a decrease in interpurchase time. These data can help the firm decide the right level of upgrades by trading off the benefits from accelerated purchases against the cost of providing an upgrade.

[Insert Figure 5 about here]

CONCLUSION

The aim of this paper is to demonstrate the value of collecting satisfaction ratings in the emerging service environment where firms can track objective service performance. The two main issues our research attempts to address via a unique data set obtained from the quick service restaurant and the auto rental industries are (1) the potential *within-individual* selection bias in satisfaction ratings and (2) the difference between customers' perceived and objective service performance. The panel nature of our data, along with ratings (or lack thereof) for all transactions, allows us to deal with the selection issue. In addition, the availability of objective service performance measures helps us circumvent the problem of using a perceived service performance as a proxy. We find evidence that customer satisfaction acts as a reasonable proxy for service performance and provides additional information value to firms as a pathway bridging the impact of objective service performance to customer purchase behavior. The fact that we find a similar pattern of results in the two very different service settings, one with a time-based measure of service quality and with a non time-based measure, suggests that these results are not idiosyncratic to setting. Overall, the implication is that firms need to continue to collect customer satisfaction data. While in some sense these results may not be unexpected, they do open up new questions for further research. First, it is possible that the customer satisfaction data provide richer perceptual feedback than has been previously assumed. In other words, customer satisfaction may capture more than just the difference between expectation and service performance as also performance inconsistency. Second, it could be that as firms have strong priors on the exact set of objective measures that map to service quality and satisfaction (as in the case of both settings here), they may be overlooking other objective measures that consumers find to be relevant.

Our analysis suffers from some limitations, primarily driven by the nature of the data. First, our data come from one firm in each of the two industries. Second, in these two industries we have a clear objective metric of service performance, which might not be easily accessible in other industries. Third, we are unable to model competitive effects. Finally, 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.

References

Anderson, E. W. and Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, 12(2):125–143.

Bickart, B. and Schmittlein, D. (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, M. J. (1992). Servicescapes: the impact of physical surroundings on customers and employees. *Journal of Marketing*, 56(2):57–71.

Bolton, R. 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, R. N., Lemon, K. N., and Bramlett, M. D. (2006). The effect of service experiences over time on a supplier's retention of business customers. *Management Science*, 52(12):1811–1823.

Boulding, W., Kalra, A., Staelin, R., and Zeithaml, V. A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, 30(1):7–27.

Bowman, D. and Narayandas, D. (2001). Managing customer-initiated contacts with manufacturers: The impact on share of category requirements and word-of-mouth behavior. *Journal of Marketing Research*, 38(3):281–297.

Carroll, W. J. and Grimes, R. C. (1995). Evolutionary change in product management: Experiences in the car rental industry. *Interfaces*, 25(5):84–104.

Chase, R. B. and Dasu, S. (2001). Want to perfect your company's service? Use behavioral science. *Harvard Business Review*, 79(6):79–84.

Churchill, G. A. and Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of Marketing Research*, 19(4):491–504.

Dellarocas, C. and Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, 21(2):277–285.

Dong, X., Janakiraman, R., and Xie, Y. (2014). The effect of survey participation on consumer behavior: The moderating role of marketing communication. *Marketing Science*, 33(4):567–585.

Ebiquity (2014). Global customer service barometer: Findings in the United States. `http://about.americanexpress.com/news/pr/2014/outstanding-service-spend-more-spread-word.aspx`.

Farris, P. W., Bendle, N. T., Pfeifer, P. E., and Reibstein, D. J. (2010). *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance*. Wharton School Publishing, Upper Saddle River, NJ, 2nd edition.

Fishbein, M. and Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., and Bryant, B. E. (1996). The American Customer Satisfaction Index: Nature, purpose, and findings. *Journal of Marketing*, 60(4):7–18.

Gijsenberg, M. J., van Heerde, H. J., and Verhoef, P. C. (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, A., Seiders, K., and Voss, G. B. (2011). Enough is enough! The fine line in executing multichannel relational communication. *Journal of Marketing*, 75(4):94–109.

Greene, W. H. (2011). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 7th edition.

Grewal, R., Chandrashekaran, M., and Citrin, A. V. (2010). Customer satisfaction heterogeneity and shareholder value. *Journal of Marketing Research*, 47(4):612–626.

Groves, R. M., Cialdini, R. B., and Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4):475–495.

Hansen, D. E. and Danaher, P. J. (1999). Inconsistent performance during the service encounter: What's a good start worth? *Journal of Service Research*, 1(3):227–235.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Hess, R. L., Ganesan, S., and Klein, N. M. (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. D., Kelley, S. W., and Chung, B. C. (2003). A CIT investigation of servicescape failures and associated recovery strategies. *Journal of Services Marketing*, 17(4):322–340.

Inside Research (2012). Special report: US CSM spend back. 23(3):16.

Jiang, L., Hoegg, J., and Dahl, D. W. (2013). Consumer reaction to unearned preferential treatment. *Journal of Consumer Research*, 40(3):412–427.

Johnson, M. D., Anderson, E. W., and Fornell, C. (1995). Rational and adaptive performance expectations in a customer satisfaction framework. *Journal of Consumer Research*, 21(4):695–707.

Keiningham, T. L., Cooil, B., Andreassen, T. W., and Aksoy, L. (2007). A longitudinal examination of Net Promoter and firm revenue growth. *Journal of Marketing*, 71(3):39–51.

Kopalle, P. K. and Lehmann, D. R. (1995). The effects of advertised and observed quality on expectations about new product quality. *Journal of Marketing Research*, 32(3):280–290.

Lund, D. J. and Marinova, D. (2014). Managing revenue across retail channels: The interplay of service performance and direct marketing. *Journal of Marketing*, 78(5):99–118.

Luo, X., Homburg, C., and Wieseke, J. (2010). Customer satisfaction, analyst stock recommendations, and firm value. *Journal of Marketing Research*, 47(6):1041–1058.

McCollough, M. A., Berry, L. L., and Yadav, M. S. (2000). An empirical investigation of customer satisfaction after service failure and recovery. *Journal of Service Research*, 3(2):121–137.

Mithas, S., Krishnan, M. S., and Fornell, C. (2005). Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing*, 69(4):201–209.

Mitra, D. and Golder, P. N. (2006). How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Science*, 25(3):230–247.

Mittal, V. and Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, 38(1):131–142.

Narayanan, S. and Manchanda, P. (2012). An empirical analysis of individual level casino gambling behavior. *Quantitative Marketing and Economics*, 10(1):27–62.

Ofir, C. and Simonson, I. (2007). The effect of stating expectations on customer satisfaction and shopping experience. *Journal of Marketing Research*, 44(1):164–174.

Oliver, R. L. (1977). Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology*, 62(4):480.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4):460–469.

Olsen, L. L. and Johnson, M. D. (2003). Service equity, satisfaction, and loyalty: From transaction-specific to cumulative evaluations. *Journal of Service Research*, 5(3):184–195.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perception of service quality. *Journal of Retailing*, 64(1):12–40.

Popuri, Y. D. and Bhat, C. R. (2003). On modeling the choice and frequency of home-based telecommuting. *Transportation Research Record*, 1858:55–60.

Rust, R. T., Inman, J. J., Jia, J., and Zahorik, A. J. (1999). What you don't know about customer-perceived quality: The role of customer expectation distributions. *Marketing Science*, 18(1):77–92.

Rust, R. T. and Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing*, 69(2):193–215.

Seiders, K., Voss, G. B., Grewal, D., and Godfrey, A. L. (2005). Do satisfied customers buy more? Examining moderating influences in a retailing context. *Journal of Marketing*, 69(4):26–43.

Sriram, S., Chintagunta, P. K., and Manchanda, P. (2015). Service quality variability and termination behavior. *Management Science*, page forthcoming.

Terza, J. V. (1987). Estimating linear models with ordinal qualitative regressors. *Journal of Econometrics*, 34(3):275–291.

Urlocker, Z. (2012). Customer satisfaction by the numbers: An industry breakdown. *Forbes*. `http://www.forbes.com/sites/ciocentral/2012/04/19/customer-satisfaction-by-the-numbers-an-industry-breakdown/`.

Verhoef, P. 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, R., Thompson, G. M., and Louviere, J. J. (1999). Configuring service operations in accordance with customer needs and preferences. *Journal of Service Research*, 1(3):262–274.

Voss, G. B., Parasuraman, A., and Grewal, D. (1998). The roles of price, performance, and expectations in determining satisfaction in service exchanges. *Journal of Marketing*, 62(4):46–61.

Weaver, D. and Brickman, P. (1974). Expectancy, feedback, and disconfirmation as independent factors in outcome satisfaction. *Journal of Personality and Social Psychology*, 30(3):420.

Ying, Y., Feinberg, F., and Wedel, M. (2006). Leveraging missing ratings to improve online recommendation systems. *Journal of Marketing Research*, 43(3):355–365.

Zhang, W. and Kalra, A. (2014). Joint examination of quality choice and satisfaction: The impact of circumstantial variables. *Journal of Marketing Research*, 51(4):448–462.

Table 1: Descriptive Statistics for Transactions (Study 1)

(a) Transactions from customers who rated at least once (n = 1,505,529)

	Mean	Std. Dev	Median	Min.	Max.
Delivery Time (hours)	0.62	0.21	0.59	0.03	2.00
Coupon Redemption	0.63	0.48	1	0	1
Purchase Amount (dollars)	22.95	9.03	21.05	0.01	99.99
Interpurchase Time (days)	36.83	45.71	19	1	361
Number of Transactions Since the Last Rating	4.49	4.70	3	1	127

(b) Transactions from customers who rated at least once and purchased 3 or more times (n = 1,257,174)

	Mean	Std. Dev	Median	Min.	Max.
Delivery Time (hours)	0.62	0.21	0.59	0.03	2.00
Coupon Redemption	0.63	0.48	1	0	1
Purchase Amount (dollars)	22.98	9	21.09	0.01	99.99
Interpurchase Time (days)	34.14	41.44	18	1	353
Number of Transactions Since the Last Rating	4.66	4.78	3	1	127

Table 2: Descriptive Statistics for Transactions with vs. without Satisfaction Rating (Study 1)

(a) Transactions with satisfaction rating (n = 22,855)

	Mean	Std. Dev	Median	Min.	Max.
Delivery Time (hours)	0.62	0.21	0.59	0.04	1.94
Coupon Redemption	0.63	0.48	1	0	1
Purchase Amount (dollars)	23.03	8.86	21.23	1.99	95.49
Interpurchase Time (days)	41.76	47.9	24	1	352
Number of Transactions Since the Last Rating	3.28	2.96	2	1	63

(b) Transactions without satisfaction rating (n = 118,446)

	Mean	Std. Dev	Median	Min.	Max.
Delivery Time (hours)	0.62	0.21	0.59	0.03	1.99
Coupon Redemption	0.63	0.48	1	0	1
Purchase Amount (dollars)	22.9	9.03	20.89	0.8	99.43
Interpurchase Time (days)	32.59	39.97	17	1	339
Number of Transactions Since the Last Rating	4.84	4.65	3	1	62

Table 3: Descriptive Statistics for Transactions from Raters vs. Non-Raters (Study 1)

(a) Transactions from customers who rated at least once (n = 1,505,529)

	Mean	Std. Dev	Median	Min.	Max.
Delivery Time (hours)	0.62	0.21	0.59	0.03	2.00
Coupon Redemption	0.63	0.48	1	0	1
Purchase Amount (dollars)	22.95	9.03	21.05	0.01	99.99
Interpurchase Time (days)	36.83	45.71	19	1	361

(b) Transactions from customers who never rated (n = 7,014,948)

	Mean	Std. Dev	Median	Min.	Max.
Delivery Time (hours)	0.62	0.22	0.59	0.03	2.00
Coupon Redemption	0.67	0.47	1	0	1
Purchase Amount (dollars)	22.31	9.61	20.14	0.01	99.99
Interpurchase Time (days)	41.05	52.97	21	1	362

Table 4: Parameter Estimates from the Satisfaction Models (Study 1)

	Null Model (1)	Proposed Model (2)
Rating Incidence		
Intercept		-0.7481
		(0.0152)***
Disconfirmation		0.0370
		(0.0260)
Performance Inconsistency		-0.0113
		(0.0617)
Coupon Redemption		0.0142
		(0.0113)
Number of Transactions Since the Last Rating		-0.0652
		(0.0016)***
Satisfaction Rating		
Intercept	3.4679	2.6952
	(0.0935)***	(0.0560)***
Disconfirmation	-0.9639	-0.5835
	(0.0822)***	(0.0524)***
Performance Inconsistency	-1.4716	-0.8969
	(0.2337)***	(0.1263)***
Coupon Redemption	0.0503	0.0182
	(0.0424)	(0.0242)
Cut-point: Rating = 2	0.3938	0.2379
	(0.0856)***	(0.0172)***
Cut-point: Rating = 3	0.9911	0.5930
	(0.0757)***	(0.0240)***
Cut-point: Rating = 4	2.0421	1.1986
	(0.0621)***	(0.0321)***
Error Correlation (ρ_{12})		0.4111
		(0.0635)***
Smoothing Factor (δ)		0.40
-log(Likelihood)	8391.4	41891.4
AIC	16798.8	83822.9
Number of Observations	21019	141301

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: The estimates of random coefficients are omitted to save space.

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

		Partial Models			
	Proposed Model	Actual Rating	Delivery Time	Disconfirmation & Inconsistency	Residuals
	(1)	(2)	(3)	(4)	(5)
Intercept	1.8076	1.9532	1.7728	1.8127	1.8133
	(0.1027)***	(0.1158)***	(0.1037)***	(0.1028)***	(0.1028)***
Predicted Satisfaction	-0.0342				
	(0.0110)***				
Actual Satisfaction		-0.0307			
		(0.0129)**			
Residuals[1]					0.0070
					(0.0103)
Delivery Time			0.0783		
			(0.0505)		
Disconfirmation	0.0406			0.0459	0.0456
	(0.0491)			(0.0490)	(0.0490)
Performance Inconsistency	0.0132			0.0097	0.0091
	(0.1215)			(0.1215)	(0.1215)
Coupon Redemption	0.1967	0.1968	0.1965	0.1968	0.1971
	(0.0223)***	(0.0223)***	(0.0223)***	(0.0223)***	(0.0223)***
Purchase Amount	0.1595	0.1593	0.1560	0.1577	0.1576
	(0.0310)***	(0.0310)***	(0.0311)***	(0.0310)***	(0.0310)***
lag (Interpurchase Time)	0.1858	0.1857	0.1858	0.1861	0.1861
	(0.0092)***	(0.0092)***	(0.0092)***	(0.0092)***	(0.0092)***
-log(Likelihood)	18424.9	18429.9	18428.9	18428.0	18431.2
AIC	36873.8	36875.7	36873.8	36876.0	36886.5
RMSE	1.4983	1.4985	1.4986	1.4986	1.4986
MAD	1.5119	1.5144	1.5143	1.5152	1.5127
MAE	1.1994	1.1996	1.1996	1.1997	1.1996
Number of Observations	21019	21019	21019	21019	21019

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: The estimates of random coefficients are omitted to save space.

[1] Residuals from Equation 2

Table 6: Parameter Estimates from the Robustness Checks (Study 1)

	Purchase Amount (1)	Product Images (2)	Across Selection (3)	Most Recent (4)	Simple Average (5)
Rating Incidence					
Disconfirmation	0.0341	0.2699	0.0423	0.0339	0.0247
	(0.0260)	(0.0523)***	(0.0387)	(0.0216)	(0.0265)
Performance Inconsistency	-0.0110	0.3150	0.0774	-0.0131	-0.0113
	(0.0617)	(0.1226)**	(0.0834)	(0.0617)	(0.0617)
Product Image Evaluations		-0.0049			
		(0.0471)			
Number of Transactions Since the Last Rating	-0.0652	-0.0797	-0.1187	-0.0652	-0.0652
	(0.0016)***	(0.0041)***	(0.0045)***	(0.0016)***	(0.0016)***
Satisfaction Rating					
Disconfirmation	-0.5883	-0.7771	-0.3760	-0.3932	-0.6033
	(0.0524)***	(0.0861)***	(0.1122)***	(0.0441)***	(0.0530)***
Performance Inconsistency	-0.9040	-0.8615	-0.7580	-0.9035	-0.9113
	(0.1263)***	(0.2089)***	(0.2514)***	(0.1257)***	(0.1261)***
Product Image Evaluations		-0.1039			
		(0.0836)			
Error Correlation (ρ_{12})	0.4111	0.4781	0.1350	0.4182	0.4166
	(0.0635)***	(0.1251)***	(0.0929)	(0.0635)***	(0.0636)***
Interpurchase Time					
Predicted Satisfaction	0.0078	-0.0435	-0.0376	-0.0344	-0.0348
	(0.0031)**	(0.0243)*	(0.0142)***	(0.0106)***	(0.0106)***
Disconfirmation	-0.0262	0.1865	0.1529	0.0953	0.0772
	(0.0141)*	(0.0831)**	(0.0994)	(0.0400)**	(0.0484)
Performance Inconsistency	0.0389	-0.5623	-0.2818	0.0294	0.0377
	(0.0341)	(0.1971)***	(0.2167)	(0.1133)	(0.1133)
Product Image Evaluations		0.1270			
		(0.0783)			
Smoothing Factor (δ)	0.40	0.40	0.40	0.40	0.40
Adjusted R-squared			0.0446		
-log(Likelihood)	2967.1	5708.1		18533.9	18535.2
AIC	5958.2	11442.3		37091.8	37094.3
Number of Observations	141301	24723	141301	141301	141301

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: The estimates of intercepts, control variables, cut-points and random coefficients are omitted to save space.

Table 7: Descriptive Statistics for Transactions (Study 2)

(a) Transactions from customers who rated at least once (n = 4,386,650)

	Mean	Std. Dev	Median	Min.	Max.
Car-class Upgrade	0.55	0.50	1	0	1
Daily Rental Price ($100)	0.37	0.21	0.33	0	1.31
Rental Duration (day)	4.14	2.93	3	1	28
Interpurchase Time (day)	42.94	68.65	18	1	897
Number of Transactions Since the Last Rating	9.82	10.81	6	1	109

(b) Transactions from customers who rated at least once and purchased 3 or more times (n = 4,088,455)

	Mean	Std. Dev	Median	Min.	Max.
Car-class Upgrade	0.56	0.50	1	0	1
Daily Rental Price ($100)	0.37	0.21	0.33	0	1.31
Rental Duration (days)	4.08	2.84	3	1	28
Interpurchase Time (days)	40.50	62.72	18	1	805
Number of Transactions Since the Last Rating	9.95	10.84	6	1	109

Table 8: Descriptive Statistics for Transactions with vs. without Satisfaction Rating (Study 2)

(a) Transactions with satisfaction rating (n = 21,681)

	Mean	Std. Dev	Median	Min.	Max.
Car-class Upgrades	0.53	0.5	1	0	1
Daily Rental Price ($100)	0.36	0.21	0.31	0	1.31
Rental Duration (days)	4.48	3.25	4	1	28
Interpurchase Time (days)	77.41	94.71	41	1	750
Number of Transactions Since the Last Rating	7.11	8.05	4	1	95

(b) Transactions without satisfaction rating (n = 246,032)

	Mean	Std. Dev	Median	Min.	Max.
Car-class Upgrades	0.56	0.5	1	0	1
Daily Rental Price ($100)	0.37	0.21	0.33	0	1.31
Rental Duration (days)	4.05	2.79	3	1	28
Interpurchase Time (days)	37.68	58.68	16	1	768
Number of Transactions Since the Last Rating	10.27	11.26	6	1	107

Table 9: Parameter Estimates from the Satisfaction and the Interpurchase Time Models (Study 2)

	Proposed Model
Rating Incidence	
Intercept	-1.4226
	(0.0178)***
Disconfirmation	-0.0198
	(0.0080)**
Performance Inconsistency	0.0019
	(0.0292)
Rental Price	-0.1517
	(0.0216)***
Number of Transactions Since the Last Rating	-0.0123
	(0.0005)***
Satisfaction Rating	
Intercept	2.8805
	(0.1012)***
Disconfirmation	0.0969
	(0.0163)***
Performance Inconsistency	-0.1826
	(0.0577)***
Rental Price	-0.4672
	(0.0515)***
Error Correlation (ρ_{12})	0.4405
	(0.0980)***
Interpurchase Time	
Intercept	2.2342
	(0.0708)***
Predicted Satisfaction	-0.0154
	(0.0058)***
Disconfirmation	-0.0168
	(0.0171)
Performance Inconsistency	0.0733
	(0.0615)
Rental Price	0.1913
	(0.0486)***
Smoothing Factor (δ)	0.10
-log(Likelihood)	19806.9
AIC	39679.9
Number of Observations	267713

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: The estimates of cut-points and random coefficients are omitted to save space.

Table 10: Consistent Results across Both Settings

(a) Quick Service Restaurant Industry

	Rating Incidence	Satisfaction Rating	Interpurchase Time
Disconfirmation		Negative	
Performance Inconsistency		Negative	
Customer Satisfaction			Negative

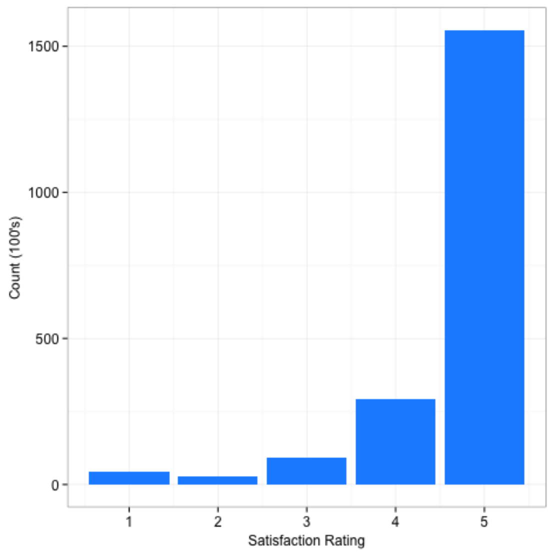
(b) Auto Rental Industry

	Rating Incidence	Satisfaction Rating	Interpurchase Time
Disconfirmation	Negative	Positive[1]	
Performance Inconsistency		Negative	
Customer Satisfaction			Negative

[1] Implies the same direction with the negative impact of disconfirmation in (a)

Figure 1: Summary of Satisfaction Ratings (Study 1)

(a) Distribution of Satisfaction Ratings



(b) Delivery Time across Satisfaction Ratings

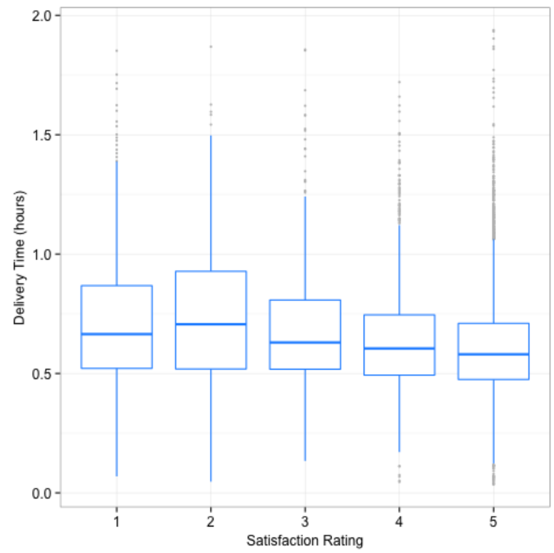
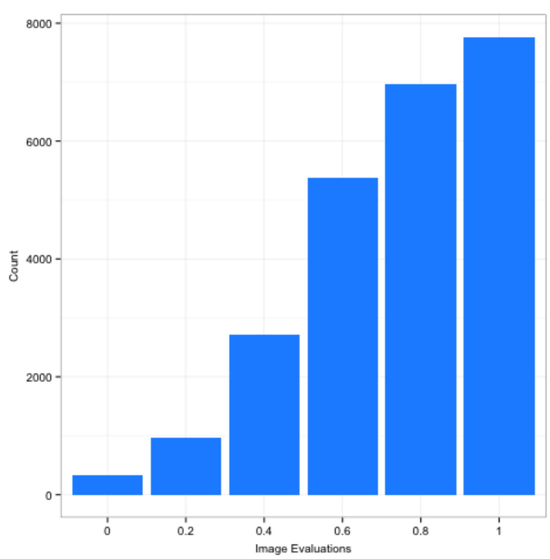


Figure 2: Summary of Average Product Image Evaluations (Study 1)

(a) Distribution of Image Evaluations



(b) Interpurchase Time across Image Evaluations

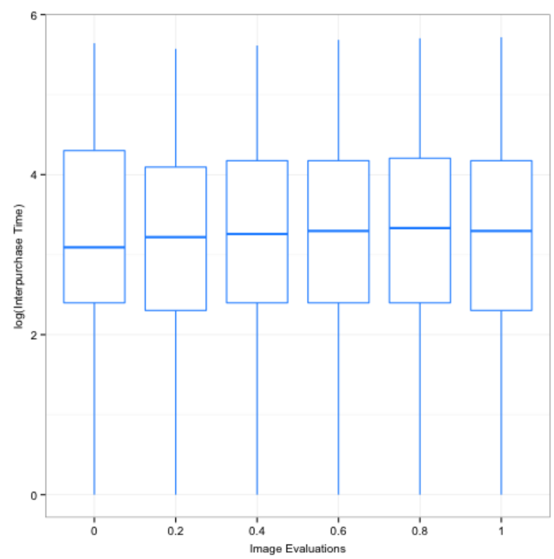
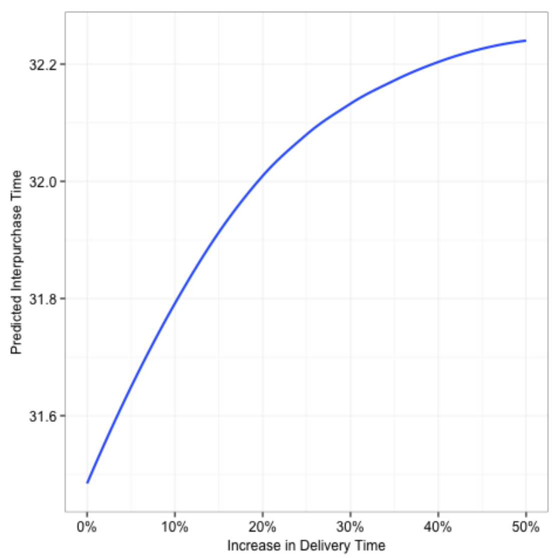


Figure 3: The Effect of Delay in Service (Study 1)

(a) Interpurchase Time



(b) Customer Satisfaction Ratings

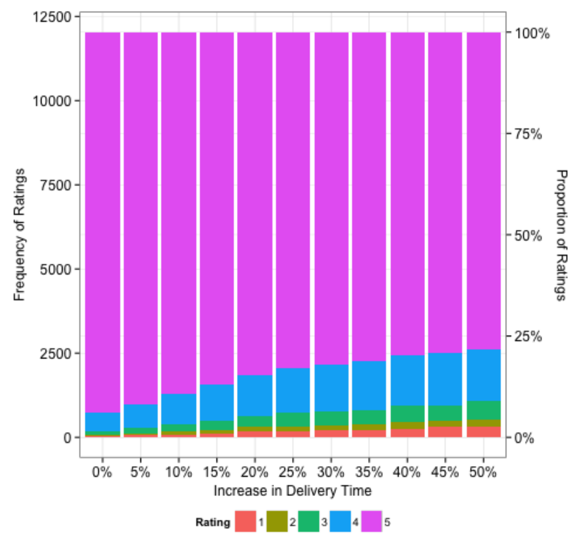


Figure 4: Distribution of Satisfaction Ratings and Free Upgrades (Study 2)

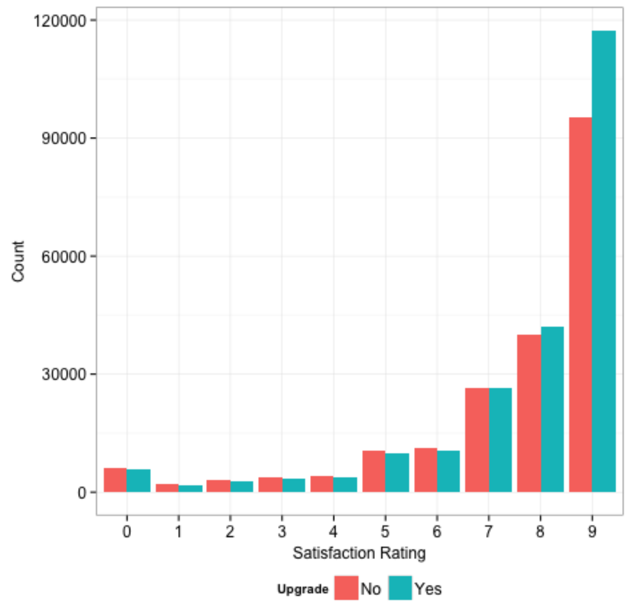


Figure 5: The Effect of Car-Class Upgrades (Study 2)

(a) Interpurchase Time

(b) Customer Satisfaction Ratings