MICHIGAN
ROSS SCHOOL OF BUSINESS

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
July 2015

The Value of Measuring Customer Satisfaction

Jihoon Cho
Anocha Aribarg
Puneet Manchanda*

Current version: July 31, 2015

*Jihoon Cho is a doctoral student, Anocha Aribarg is Associate Professor of Marketing and Puneet Manchanda is Isadore and Leon Winkelman Professor and Professor of Marketing, all at the Stephen M. Ross School of Business at the University of Michigan. The authors are grateful to Xiaojing Dong, Fred Feinberg, Yesim Orhun, Adithya Pattabhiramaiah, S. Sriram, and Ying Xie, and seminar participants at the University of Michigan and the University of Cambridge, INFORMS Marketing Science Conference 2013, UT Dallas FORMS conference 2015 for helpful comments. The authors are also grateful to anonymous companies and the Wharton Customer Analytics Initiative for providing the data, as well as Melissa Hartz and Ben Adams for answering our questions about the auto rental industry. 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*.

Abstract

Firms in multiple industries have invested significantly in measuring customer satisfaction. These firms use customer satisfaction as an indicator of their service performance, as objective measures of service performance are usually hard to obtain. To justify the investment, firms are interested in establishing the link between customer satisfaction and desirable customer behavior such as repurchase. Facing the same constraints as firms, most academic research relies on survey-based self-reported measures of perceived service performance, customer satisfaction and purchase intent. This prevalent method suffers from two limitations. First, perceived and actual service performance tend to differ due to a variety of factors. Second, not all customers reply to surveys, opening up the possibility of the results being biased by selection.

This paper assesses the value of measuring customer satisfaction when objective service performance data are available. Unlike most previous research that uses cross-sectional self-reported 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 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, 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.

Using these two novel data sets, we model rating incidence, satisfaction rating and interpurchase time using a simultaneous equations model. Our results indicate the presence of within-individual selection. We also find that satisfaction ratings reflect objective service performance. Interestingly, despite its weak direct effect, we find a strong indirect effect of objective service performance on interpurchase time, operating through customer satisfaction. Such evidence sheds light on the value customer satisfaction as a pathway bridging objective service performance to customer purchase behavior.

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

1 Introduction

Firms in multiple industries have invested significantly in measuring and tracking customer satisfaction. As of 2012, the 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, as 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 (Forbes 2012). In general, to justify the investment in the measurement effort, firms are interested in establishing a clear link between customer satisfaction and desirable customer behavior. However, it is not clear whether customer satisfaction mirrors firms' objective performance and whether the relationship between unobservable constructs such as customer satisfaction and observable behavior such as purchasing is substantial (Gupta and Zeithaml 2006). In this research, we aim to assess the value of measuring customer satisfaction by asking two specific questions - does customer satisfaction reflect firms' objective service performance and does customer satisfaction provide additional information relative to what can be obtained from objective service performance?

The academic literature (e.g., Oliver 1980; Bolton and Drew 1991) 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-report measures of perceived service performance,[1] customer satisfaction and purchase intent. 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 expec-

[1]Notable exceptions are discussed in Section §2.3

tations 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; Bolton and Drew 1991), understanding how objective service performance is linked to customers' desirable behavior can provide firms with more actionable guidelines for improvement (e.g., Grewal, Chandrashekaran, and Citrin 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).

This paper assesses the value of measuring customer satisfaction when objective service performance data are available. Unlike most previous 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 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 drivers of satisfaction rating. In the second stage, we model customers' interpurchase time as a function of the predicted customer satisfaction rating along with other controls. Our results show that the decision to rate and the satisfaction rating are correlated, providing support for within-individual self-selection phenomenon. We also 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 disconfirmation and performance inconsistency direct effects, customer satisfaction has a strong effect on interpurchase time. 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 our results to show that failing to maintain good service performance leads to undesirable changes in customer purchasing behavior.

The remainder of the paper is organized as follows. In §2, we present an overview of the theoretical and methodological issues of past studies in service performance and customer satisfaction. In §3 we present our model and estimation procedure. In §4 and §5, we 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 in §4 and §5. Finally, we conclude with a discussion of our key findings, research limitations, and directions for future research in §6.

2 Conceptual Background

2.1 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 typically references the customer's satisfaction with a specific, discrete service encounter (Bitner and Hubert 1994). This transaction-specific customer satisfaction has been measured by asking survey participants to consider the last experience they had (e.g., Olsen and Johnson 2003; Agustin and Singh 2005). The latter refers to the customer's cumulative evaluation based on all encounters and experiences with the organization. Notably, the attitude-based satisfaction resembles the concept of perceived service quality in the service literature (Parasuraman, Zeithaml, and Berry 1985, 1988). In this research, we focus on transaction-based satisfaction.

Early satisfaction research identifies two important constructs that influence customer satisfaction: performance expectation and disconfirmation (Ilgen 1971). The expectation construct is motivated by Helson (1948)'s adaptation level theory which posits that individuals perceive stimuli only in relation to an adapted standard, which is a function of the stimuli themselves, characteristics of the individuals, and the context. Once the adaptation level or expectation is established, it creates a "latitude of acceptance" where positive or negative deviation of perceived performance within a close vicinity of the expectation leads to the assimilation of perceived performance to the expectation (Sherif and Hovland 1961). Early studies manipulate expectation using consumption description (Olshavsky and Miller 1972) or ad-like consumption information (Anderson 1973). Disconfirmation is the discrepancy between performance and expectation. Disconfirmation has been viewed as a distinct and independent construct from expectation, and the two are assumed to have additive effects on customer satisfaction (Ilgen 1971). Positive disconfirmation (when performance exceeds expectation) is hypothesized to increase customer satisfaction and negative disconfirmation to decrease customer satisfaction (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, Zeithaml, and Berry (1988) propose a multi-item scale called SERVQUAL to measure perceived service performance where disconfirmation is derived as a difference score between perceived performance and performance expectation ratings on different service aspects. To 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-report 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, 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, Parasuraman, and Grewal 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 (Hogarth and Einhorn 1992) and derive it as a function of the objective performance that varies over time. This specification is consistent with how previous research uses this process to conceptualize how adaptation level or expectation gets adjusted over time (Boulding et al. 1993). We then conceptualize that disconfirmation - the difference between the derived expectation and objective performance - ultimately affects transaction-based satisfaction.

2.2 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., Voss, Godfrey, and Seiders 2010; Godfrey, Seiders, and Voss 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, Seiders, and Voss (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, Krishnan, and Fornell (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 *within-individual* selection process in satisfaction rating. In our study, we obtain a unique dataset where we observe satisfaction ratings and purchases over multiple time periods. Individuals in our dataset 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. Our treatment of selection is more in line with (Ying, Feinberg, and Wedel 2006). That is, we model the rating incidence and the actual satisfaction rating simultaneously and address the *within-individual* selection bias in customer satisfaction rating.

2.3 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., Anderson and Sullivan 1993; Churchill and Surprenant 1982; Oliver 1980). We define objective performance as an unbiased measure of a service or product. Objective measures of performance can be based on either observable and concrete metrics (e.g., minutes, number of defects) or expert ratings (Lichtenstein and Burton 1989; 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 (Bitner and Hubert 1994), as well as prior expectations of performance (Anderson and Sullivan 1993).

Several studies have managed to secure objective performance measures, typically in contractual service settings. Bolton and Drew (1991) examine how different objective performance measures - customer complaints about a local telephone service (e.g., difficulty with voice transmission and/or call connection) - affect overall perceived performance and, subsequently, how the perceived performance affects customer satisfaction. Bolton and Lemon (1999) obtain objective measures of customers' phone usage and payment but use them as control variables. Bolton, Lemon, and Bramlett (2006) link the objective performance of a supplier's engineering service (e.g., work minutes per a support request), and Sriram, Chintagunta, and Manchanda (2015) link signal quality of a video-on-demand service to customer retention. However, these papers do not study satisfaction in the same framework. Grewal, Chandrashekaran, and Citrin (2010) find that objective performance at the aggregate level (e.g., percentage of on-time arrival, mishandled baggage and complaint) to significantly affect the overall attitude-based customer satisfaction in the airline industry. Finally, Gijsenberg, van Heerde, and Verhoef (2015) find the proportion of successful connections in the railway service industry to affect aggregate customer satisfaction, and also document the asymmetrically larger effect of failed than successful service performance.

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.

2.4 Impact of Satisfaction on Purchase Behavior

Satisfaction research has consistently shown the impact of satisfaction on purchase intention (Anderson and Sullivan 1993; Oliver 1980; Seiders et al. 2005) and on downstream business outcomes such as service usage (Bolton and Lemon 1999), retention (Rust and Zahorik 1993), share of customer wallet (Bowman and Narayandas 2001), and firm's financial performance (Luo, Homburg, and Wieseke 2010). Nonetheless, some research finds the direct main effect of satisfaction on individual-level purchase behavior (Seiders et al. 2005; Verhoef 2003) 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. If researchers can match participants' survey answers to their purchase records in a customer database, only a survey at T_0 is necessary (Mittal and Kamakura 2001; Seiders et al. 2005). Sometimes researchers also administered a survey to capture the customer's actual 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). Other studies collect data at two time points to measure satisfaction at one time point and some other variables of interest at the other, such as service duration (Bolton 1998) and usage intensity (Bolton and Lemon 1999). Bolton and Drew (1991) attempted to collect satisfaction ratings along with other variables at three time points to examine changes in customers' satisfaction in response to a phone company's network upgrade program. All this research encounters the usual problem of decreasing 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.

2.5 Performance Consistency

The service management literature has emphasized the importance of consistent service performance in maintaining high perceived quality, service value and customer loyalty (Hart, Heskett, and Sasser Jr. 1989). Reliability - partly described as consistency in performance - is found to be an important determinant of overall perceived service performance (Parasuraman, Zeithaml, and Berry 1988). For example, McCollough, Berry, and Yadav (2000) conduct scenario-based experiments to show that customer satisfaction is lower after service failure and recovery (even with high-recovery performance) than in the case of consistent error-free service. 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.

At the same time, other research provides some caveats that some particular situations of performance inconsistency may also lead to positive outcomes. Bolton, Lemon, and Bramlett (2006) find a few extremely favorable experiences to be critical for business customers' subsequent repatronage behavior such as system support contract renewal. Similarly, Verhoef, Antonides, and de Hoog (2004) find the peak and Hansen and Danaher (1999) find the end experiences within a service encounter to increase satisfaction. In business-to-business logistics service setting, Van Doorn and Verhoef (2008) show that crises - unusually negative performance - may help rejuvenate long-term business relationships and thus positively affect customer share. Sriram, Chintagunta, and Manchanda (2015) show that high levels of service variability can increase customer retention when the general service performance is low.

To measure performance inconsistency, previous literature in service performance use the number of extremely positive or negative performance (Bolton, Lemon, and Bramlett 2006; Van Doorn and Verhoef 2008), the proportion of successful performance (Gijsenberg, van Heerde, and Verhoef 2015), the magnitude of service failure (Smith, Bolton, and Wagner 1999), and the variance of service performance (Meyer 1981; Sriram, Chintagunta, and Manchanda 2015). Taking advantage of multi-period panel data of objective performance, this paper specifies the evolution of performance inconsistency by calculating 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.

3 Modeling Approach

3.1 Model

In this section, we model satisfaction rating incidence, satisfaction rating, and interpurchase time as three separable but related processes, by constructing a set of simultaneous equations (e.g., Lee, Maddala, and Trost 1980) 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, respectively. 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). To do so, we specify a correlated error structure for the binary Probit and the ordered Probit models. 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 are specified as follows:

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

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

$$INT_{ijt} = \gamma_1 X + \gamma_2 \widehat{SAT}_{ijt-1} + \epsilon^{INT}_{ijt} \tag{3}$$

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} are observed rating incidence, satisfaction ratings, and log-transformed interpurchase time in the data, respectively. Z^{INC}, Z^{SAT}, and X represent sets of explanatory variables for rating incidence, sat-

[2]Alternatively, the probability of purchase given that the customer has not purchased since the previous service encounter could be modeled as a hazard function (e.g., Gupta 1991). We chose a linear model of log-transformed interpurchase time because of its simplicity.

isfaction rating, and interpurchase time, respectively. $\widehat{SAT}_{ijt}$ is the predicted customer satisfaction generated from Equation 1 and 2.

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 K observed ordinal scale points. For identification purposes, we set the lowest and the highest cutoffs, μ_0 and μ_5, to $-\infty$ and ∞. In addition, the second lowest cutoff, μ_1, is fixed to zero (e.g., Ying, Feinberg, and Wedel 2006). 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^{INC}_{ijt} and ϵ^{SAT}_{ijt} be unity.

Following the previous studies (Ilgen 1971; Olshavsky and Miller 1972; 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, where the expectation is updated based on service performance and prior expectation in the previous period (e.g., Anderson and Salisbury 2003; Boulding et al. 1993; Hogarth and Einhorn 1992).

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

$PERF_{ijt}$ 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.[3] Note that, due to the availability of objective performance data, we do not need to operationalize *perceived* disconfirmation as has typically measured on a scale with end labels of "better than expected" and "worse than expected" to a question asking how close the product came to prior expectation (e.g., Oliver 1980). These self-report measures of expectation may not match customers' pre-consumption expectations due to cognitive dissonance, assimilation, or contrast (Oliver 1977). Service performance inconsistency, VAR_{ijt}, is operationalized as a cumulative standard deviation of delivery time up to the

[3]To determine the smoothing factor δ, we perform a grid search method by searching the interval from zero to one with a step size of 0.1.

current service encounter (e.g., Meyer 1981; Sriram, Chintagunta, and Manchanda 2015).[4]

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 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, Cialdini, and Couper 1992).

To control for unobserved individual characteristics that may otherwise lead to spurious within-individual correlation, a random coefficient model specification is used. Specifically, we model 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 in 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.

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.

3.2 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

[4]The first two observations of each individual are not included in the model estimation but used to calculate the cumulative standard deviation of objective service performance.

bias. We estimate the two equations simultaneously, using a partial maximum likelihood approach (Popuri and Bhat 2003).[5] 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). Using the transaction data both with and without ratings, we extend the log-likelihood function for the binary Probit model as follows:

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

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, Feinberg, and Wedel 2006).

To capture customer heterogeneity in disconfirmation and performance inconsistency, we estimate a random effects model, assuming that the parameters to be estimated across different individuals are not identical, but follow some distributions. The random effects model is estimated by a simulated maximum likelihood approach (Greene 2011). 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

[5] Constructing Inverse Mills ratio from an estimate of the Probit 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}.

unobserved variable u_{ir}:

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

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.[6]

4 Study 1: Quick Service Restaurant Industry

4.1 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, Thompson, and Louviere 1999). Timely service has been widely accepted as a key to success in the service industry because it is the first interaction in the sequence of experiences that customers have with the firm (e.g., Bitner 1992). The nature and order of these experiences thus can have an impact on overall

[6] See Terza (1987) for more detailed information on the derivation of Equation 7.

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.[7] Based on both the previous literature and our specific setting, we use delivery time as the key objective service performance measure in our analysis.

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.[8] The survey consists of six questions as below:

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

[7]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.

[8]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 (e.g., Zhang and Kalra 2014). Thus, the focus in this study is transaction-based, rather than attitude-based, satisfaction.

4.2 Data Description

4.2.1 Transaction Data

The data span a total of 1,505,529 delivery orders from 362,672 unique customers (households)[9] who provided satisfaction ratings at least once during the sample period at 55 stores in Texas and Virginia from January to December 2011.[10] 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.

[Table 1 about here.]

4.2.2 Satisfaction Ratings

From the six questions on the online survey, we use Q1 ("How likely are you to recommend us to your family and friends?") as the measure of transaction-specific customer satisfaction. Previous

[9] 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. This remains a limitation of our data.

[10] 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 §4.4.3.

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 closed to 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 (e.g., Bolton and Drew 1991).[11] 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).

[Figure 1 about here.]

Customer satisfaction ratings are provided in 2.8% of the total transactions. Considering the transactions from "raters" who provided satisfaction ratings at least once during the sample period, however, customers rate 1.73 times on average and the rating incidence is 16.0%. The low rating incidence is explained by the company's managers as due to customers' lack of knowledge about their ability to track their orders and provide feedback on the company's website.

4.2.3 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 the relevant variables - delivery time, coupon redemption, purchase amount, and interpurchase time. Table 2 presents the key variables (1) from the sample with satisfaction ratings and (2) from transactions without satisfaction ratings.

[11] Over 90% of the responses to Q2 and Q5 are missing mostly because phone-order/ carry-out customers are not likely to access to the company's website to track their order status although they are still able to access the "order tracker." A factor analysis using orthogonal Varimax rotation suggests responses Q1 load on a different factor from responses to Q3, Q4, and Q6.

[Table 2 about here.]

As can be seen from Tables 2, there is no significant difference in disconfirmation and performance inconsistency across the three different samples. While interpurchase time is somewhat longer for transactions with satisfaction ratings relative to those without ratings (41.8 vs. 32.6 days), the difference is 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-individuals selection to detect its presence and compare its magnitude to that of the within-individual selection (See §4.4.3).

[Table 3 about here.]

4.3 Results

4.3.1 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.

[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 previous literature that proposes a relationship between disconfirmation and customer satisfaction (e.g., Oliver 1980).[12] Models 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 with 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 recent 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.

4.3.2 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 along with disconfirmation and performance inconsistency to interpurchase time using only those observations for which we have satisfaction ratings.[13] We then compare our proposed model with a series of alternative models in order to answer our research question. Table 5 reports the results.

[Table 5 about here.]

[12] We also find the asymmetric effect of disconfirmation on customer satisfaction. Following Anderson and Sullivan (1993), we test 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. $\beta_{negative} = 0.0061$). The complete results are available upon request.

[13] Due to data limitations, we cannot account for competitive activity. It is important to note that the company also does not have access to competitor data.

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.

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 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 §4.3.1). 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 Model 2 provides the best fit. Thus it may seem that purely from a fit point of view, the only measure of performance needed is the actual satisfaction score. But the results from the previous section (§4.3.1) suggest that selection is likely to induce biased satisfaction ratings. In addition, if the firms use this metric alone, the only course of action available to them is to "increase the satisfaction score," with no guidance as how to do so. In contrast, our proposed model is more useful for firms as it can help link objective service performance to customers' repurchase behavior through customer satisfaction, and therefore provide implementable suggestion of how firms can improve their service.

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 mean absolute error (MAE) and root mean square error (RMSE). The results show that Model 2 has the best predictive performance with respect to MAE. However, our proposed model performs best based on RMSE, as this statistic penalizes observations that deviate further from the mean more than does MAE.

Collectively, the results highlight the value of collecting both objective and subjective measures of performance. The combined set of measure helps to assess customer behavior relevant to managerial outcomes. In addition, these metrics provide diagnostics on the impact of managerial actions on firm performance (see §4.5).

4.4 Robustness Checks

In this section, we conduct 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.

4.4.1 Customer Purchase Amount

Previous research has documented that increased satisfaction can lead to higher consumption e.g., Bolton and Lemon (1999) find that customer satisfaction increases usage of cell phone service. We test to see if our results are robust to this 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 of which dependent variable is log-transformed purchase amount. The results are reported in Table 6. 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.

[Table 6 about here.]

4.4.2 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.[14] 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.

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."[15] 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.[16] The product image evaluations during the sample period are reasonably high and they do not appear to be correlated with interpurchase time (the correlation coefficient is -0.03).

[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

[14] The proportion of franchise stores in the sample of the stores with the camera - about 62% - is broadly consistent with that in the entire sample.

[15] The raters are allowed also to give a "can't rate" rating when they cannot decide between a "good" or a "bad" rating. 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. For example, 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.

[16] As each store gets five image evaluations a day, the average image ratings per day are clustered on 0.2, 0.4, 0.6, 0.8, and 1.0.

this analysis the store-level daily average of the product image scores (from five randomly selected photos) is used as the measure of product performance for each store for a given day. Notably, 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 7, the results demonstrates yet 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 also do not find product performance to significantly affect either satisfaction or interpurchase time.

[Table 7 about here.]

4.4.3 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-individuals* 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 8 reports the results of the model estimated on this sample.

The results show that "across-individuals 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.

[Table 8 about here.]

4.4.4 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 9 report that the effect of disconfirmation is robust across different measures of customer expectation. As shown in Column (1) and (2), no direct impacts of disconfirmation are observed with respect to rating incidence and inter-purchase time. Positive disconfirmation (i.e., worse-than-expected service performance) only decreases satisfaction rating.

[Table 9 about here.]

4.5 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 leads to loss in 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 5% 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%. 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. As shown in Figure 3b, 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.

[Figure 3 about here.]

5 Study 2: Auto Rental Industry

5.1 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.[17] The previous service literature identifies free upgrades as one of the (compensatory) recovery strategies to address service failure (e.g., Hoffman, Kelley, and Chung 2003), which potentially increase customers' positive reactions. Some consumer behavior studies (e.g., Jiang, Hoegg, and Dahl 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.[18] 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.

[17]This practice is analogous to planned overbooking over multiple compartments by airlines, where economy passengers who cannot be accommodated in the coach compartment get free upgrades to business-class. However, the practice is more prevalent in the auto rental industry because there are many more types of inventory and the capacities are more evenly balanced across the different car types (Talluri and Van Ryzin 2006).

[18]The vast majority of customers submitted via the email link, but some customers submitted their survey using a touch-tone phone of which questions and possible responses are identical to those from the email link.

5.2 Data Description

5.2.1 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.[19] Across 689 airport and 2,759 non-airport locations, 1 to 68 different car-classes are offered, while 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. The 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[20], check-in/check-out date, base rental price, and satisfaction ratings. Consistent with Study 1, we derive expectation, disconfirmation and performance inconsistency from objective service performance (i.e., whether a free car-class upgrade is offered or not). To use the cumulative standard deviation of the objective service performance as a proxy of performance inconsistency, we focus on customers who rented three or more times during 2-year sample period. This limits the sample size to 4,088,455 transactions from 304,516 unique loyalty club members. As can be seen from Table 10, we do not observe substantial differences in the behavior between the sample of customers who purchased three or more times and the complete sample.

[Table 10 about here.]

5.2.2 Satisfaction Ratings

From the eight questions in the survey, 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.

[19] From the original data that contain 6,283,105 observations we take out the transactions with invalid loyalty member ID and missing car-class information. We also exclude the outliers (> 99th percentile) of rental duration, reservation date, daily rental price, and purchase frequency.

[20] We identify unique customers by a combination of each customer's loyalty club membership ID and birth date on her driver's license. By doing this, we rule out the possibility that the purchase history under a single membership consists of multiple customers (i.e., drivers).

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.[21] 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.

[Figure 4 about here.]

5.2.3 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 11 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.

[Table 11 about here.]

From Table 11, 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 a longer interpurchase time, compared to those without ratings (77.4 vs. 37.7 days). The difference between the two samples opens 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

[21] While we acknowledge that Q1 and Q3 are possibly measure overall service performance, we do not use customer responses to both questions due to the following issues in the data collection process. First, we observe a high proportion of raters (37.5%) gave the company the lowest rating for Q1, something that we find implausible and inconsistent with all the other measures. So 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. A factor analysis using orthogonal Varimax rotation suggests responses Q2 load on a different factor from the responses to Q4, Q5, Q6, Q7, and Q8.

at least once during the 24-month sample period. We therefore cannot examine across-individuals selection.

5.3 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 12 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.

[Table 12 about here.]

As shown in Table 12, 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 13 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 13 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.

[Table 13 about here.]

6 Concluding remarks

The aim of this paper is to demonstrate the value of collecting satisfaction ratings that provide information over and above that can be found in objective service performance. 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 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.

Our analysis suffers from some limitations, primarily driven by the nature of the data. First, our data come from one firm in one industry. Second, in these two industries we have a clear objective metric of service performance, which might not be available 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. We hope that future research can address these limitations.

References

Agustin, Clara, Jagdip Singh. 2005. Curvilinear effects of consumer loyalty determinants in relational exchanges. *Journal of Marketing Research* **42**(1) 96–108.

Anderson, Eugene W, Linda Court Salisbury. 2003. The formation of market-level expectations and its covariates. *Journal of Consumer Research* **30**(1) 115–124.

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

Anderson, Rolph E. 1973. Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of marketing research* 38–44.

Bickart, Barbara, David Schmittlein. 1999. The distribution of survey contact and participation in the united states: Constructing a survey-based estimate. *Journal of Marketing Research* 286–294.

Bitner, Mary Jo. 1992. Servicescapes: the impact of physical surroundings on customers and employees. *The Journal of Marketing* 57–71.

Bitner, Mary Jo, A.R. Hubert. 1994. Encounter satisfaction versus overall satisfaction versus quality. Rust T. Rust, Richard L. Oliver, eds., *Service Quality: New Directions in Theory and Practice*. Sage Publications, 72–94.

Bolton, Ruth N. 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science* **17**(1) 45–65.

Bolton, Ruth N., James H. Drew. 1991. A multistage model of customers' assessments of service quality and value. *Journal of Consumer Research* **17**(4) 375–384.

Bolton, Ruth N., Katherine N. Lemon. 1999. A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. *Journal of Marketing Research* **36**(2) 171–186.

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

Boulding, William, Ajay Kalra, Richard Staelin, Valarie A. Zeithaml. 1993. A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research* **30**(1) 7–27.

Bowman, Douglas, Das Narayandas. 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, William J, Richard C Grimes. 1995. Evolutionary change in product management: Experiences in the car rental industry. *Interfaces* **25**(5) 84–104.

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

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

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

Ebiquity. 2014. Global customer service barometer: Findings in the united states.

Farris, Paul W., Neil T. Bendle, Phillip E. Pfeifer, David J. Reibstein. 2010. *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance*. 2nd ed. Wharton School Publishing, Upper Saddle River, NJ.

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

Forbes. 2012. Customer satisfaction by the numbers: An industry breakdown.

Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, Barbara Everitt Bryant. 1996. The american customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing* **60**(4) 7–18.

Gijsenberg, Maarten J, Harald J van Heerde, Peter C Verhoef. 2015. Losses loom longer than gains: Modeling the impact of service crises on perceived service quality over time. *Journal of Marketing Research* .

Godfrey, Andrea, Kathleen Seiders, Glenn B Voss. 2011. Enough is enough! the fine line in executing multichannel relational communication. *Journal of Marketing* **75**(4) 94–109.

Grainer, Marc, Charles H. Noble, Mary Jo Bitner, Scott M. Broetzmann. 2014. What unhappy customers want. *MIT Sloan Management Review* **49**(4) 41–50.

Greene, William H. 2011. *Econometric Analysis*. 7th ed. Prentice Hall, Upper Saddle River, NJ.

Greene, William H., David A. Hensher. 2010. Ordered choices and heterogeneity in attribute processing. *Journal of Transport Economics and Policy* **44**(3) 331–364.

Grewal, Rajdeep, Murali Chandrashekaran, Alka V. Citrin. 2010. Customer satisfaction heterogeneity and shareholder value. *Journal of Marketing Research* **47**(4) 612–626.

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

Gupta, Sunil. 1991. Stochastic models of interpurchase time with time-dependent covariates. *Journal of Marketing Research* **28**(1) 1–15.

Gupta, Sunil, Valarie Zeithaml. 2006. Customer metrics and their impact on financial performance. *Marketing Science* **25**(6) 718–739.

Hart, Christopher W., James L. Heskett, W. Earl Sasser Jr. 1989. The profitable art of service recovery. *Harvard Business Review* **68**(4) 148–156.

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

Helson, Harry. 1948. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review* **55**(6) 297.

Hoffman, K Douglas, Scott W Kelley, Beth C Chung. 2003. A cit investigation of servicescape failures and associated recovery strategies. *Journal of Services Marketing* **17**(4) 322–340.

Hogarth, Robin M., Hillel J. Einhorn. 1992. Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology* **24**(1) 1–55.

Ilgen, Daniel R. 1971. Satisfaction with performance as a function of the initial level of expected performance and the deviation from expectations. *Organizational Behavior and Human Performance* **6**(3) 345–361.

Jiang, Lan, JoAndrea Hoegg, Darren W Dahl. 2013. Consumer reaction to unearned preferential treatment. *Journal of Consumer Research* **40**(3) 412–427.

Keiningham, Timothy L, Bruce Cooil, Tor Wallin Andreassen, Lerzan Aksoy. 2007. A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing* **71**(3) 39–51.

Lee, Lung-Fei, G. S. Maddala, R. P. Trost. 1980. Asymptotic covariance matrices of two-stage probit and two-stage tobit methods for simultaneous equations models with selectivity. *Econometrica* **48**(2) 491–503.

Lichtenstein, Donald R., Scott Burton. 1989. The relationship between perceived and objective price-quality. *Journal of Marketing Research* **26**(4) 429–443.

Luo, Xueming, Christian Homburg, Jan Wieseke. 2010. Customer satisfaction, analyst stock recommendations, and firm value. *Journal of Marketing Research* **47**(6) 1041–1058.

Manchanda, Puneet, Asim Ansari, Sunil Gupta. 1999. The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Science* **18**(2) 95–114.

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

Meyer, Robert J. 1981. A model of multiattribute judgments under attribute uncertainty and informational constraint. *Journal of Marketing Research* 428–441.

Mithas, Sunil, Mayuram S Krishnan, Claes Fornell. 2005. Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing* **69**(4) 201–209.

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

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

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

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

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

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

Olshavsky, Richard W, John A Miller. 1972. Consumer expectations, product performance, and perceived product quality. *Journal of marketing research* 19–21.

Parasuraman, A., Valarie A. Zeithaml, Leonard L. Berry. 1985. A conceptual model of service quality and its implications for future research. *Journal of Marketing* **49**(4) 41–50.

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

Popuri, Yasasvi D, Chandra R Bhat. 2003. On modeling the choice and frequency of home-based telecommuting. *Transportation Research Record* **1858**.

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

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

Seiders, Kathleen, Glenn B. Voss, Dhruv Grewal, Andrea L. Godfrey. 2005. Do satisfied customers buy more? examining moderating influences in a retailing context. *Journal of Marketing* **69**(4) 26–43.

Sherif, Muzafer, Carl I. Hovland. 1961. *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. Yale University Press, Oxford, England.

Smith, Amy K, Ruth N Bolton, Janet Wagner. 1999. A model of customer satisfaction with service encounters involving failure and recovery. *Journal of marketing research* 356–372.

Sriram, S., Pradeep K. Chintagunta, Puneet Manchanda. 2015. Service quality variability and termination behavior. *Management Science, Forthcoming* .

Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.

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

Van Doorn, Jenny, Peter C Verhoef. 2008. Critical incidents and the impact of satisfaction on customer share. *Journal of Marketing* **72**(4) 123–142.

Verhoef, Peter C. 2003. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing* **67**(4) 30–45.

Verhoef, Peter C., Gerrit Antonides, Arnoud N. de Hoog. 2004. Service encounters as a sequence of events: The importance of peak experiences. *Journal of Service Research* **7**(1) 53–64.

Verma, Rohit, Gary M. Thompson, Jordan J. Louviere. 1999. Configuring service operations in accordance with customer needs and preferences. *Journal of Service Research* **1**(3) 262–274.

Voss, Glenn B, Andrea Godfrey, Kathleen Seiders. 2010. How complementarity and substitution alter the customer satisfaction-repurchase link. *Journal of Marketing* **74**(6) 111–127.

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

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

Ying, Yuanping, Fred Feinberg, Michel Wedel. 2006. Leveraging missing ratings to improve online recommendation systems. *Journal of Marketing Research* **43**(3) 355–365.

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

Table 1: Descriptive Statistics: All Raters vs. Raters who Purchased 3 or More Times (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 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

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

(a) Transactions with satisfaction ratings (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

(b) Transactions without satisfaction ratings (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

Table 3: Descriptive Statistics: 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: Selection 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 Ratings		
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: Interpurchase Time Model (Study 1)

	Proposed Model	Partial Models			
		Actual Rating	Delivery Time	Disconfirmation & Inconsistency	Residuals
	(1)	(2)	(3)	(4)	(5)
Intercept	1.5993	1.9532	1.5681	1.6041	1.6039
	(0.0979)***	(0.1158)***	(0.0990)***	(0.0980)***	(0.0980)***
Predicted Satisfaction	-0.0349				
	(0.0106)***				
Actual Satisfaction		-0.0307			
		(0.0129)**			
Residuals					0.0043
					(0.0103)
Delivery Time			0.0867		
			(0.0494)*		
Disconfirmation	0.0632			0.0631	0.0632
	(0.0477)			(0.0477)	(0.0477)
Performance Inconsistency	0.0351			0.0361	0.0358
	(0.1133)			(0.1133)	(0.1134)
Coupon Redemption	0.1947	0.1968	0.1943	0.1946	0.1947
	(0.0214)***	(0.0223)***	(0.0214)***	(0.0214)***	(0.0214)***
log (Purchase Amount)	0.1447	0.1594	0.1390	0.1427	0.1427
	(0.0295)***	(0.0310)***	(0.0296)***	(0.0295)***	(0.0295)***
log (Interpurchase Time)	0.2531	0.1856	0.2533	0.2535	0.2536
	(0.0091)***	(0.0092)***	(0.0091)***	(0.0091)***	(0.0091)***
Smoothing Factor (δ)	0.40			0.40	0.40
-log(likelihood)	18535.5	18429.9	18535.4	18537.3	18540.9
AIC[2]	37095.0	36875.7	37086.8	37094.7	37105.8
MAE	1.2795	1.2793	1.2799	1.2798	1.2798
RMSE	1.5943	1.5944	1.5952	1.5952	1.5952
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 satisfaction ratings, estimated from Equation 1 and 2

[2] AIC for our proposed model (Column (1)) is reduced to 37082.1 when the insignificant variables are excluded from the equation.

Table 6: Parameter Estimates: Purchase Amount Model (Study 1)

	Proposed Model	Partial Models			
		Actual Rating	Delivery Time	Disconfirmation & Inconsistency	Residuals
	(1)	(2)	(3)	(4)	(5)
Intercept	1.4506	1.3993	1.4459	1.4496	1.4496
	(0.0270)***	(0.0308)***	(0.0273)***	(0.0271)***	(0.0271)***
Predicted Satisfaction	0.0074				
	(0.0029)**				
Actual Satisfaction		0.0124			
		(0.0035)***			
Residuals					-0.0004
					(0.0029)
Delivery Time			0.0182		
			(0.0137)		
Disconfirmation	-0.0272			-0.0272	-0.0272
	(0.0132)**			(0.0132)**	(0.0132)**
Performance Inconsistency	0.0290			0.0288	0.0288
	(0.0313)			(0.0313)	(0.0313)
Coupon Redemption	0.0101	0.0093	0.0097	0.0101	0.0101
	(0.0059)*	(0.0059)	(0.0059)	(0.0059)*	(0.0059)*
log (Purchase Amount)	0.5207	0.5202	0.5201	0.5211	0.5211
	(0.0082)***	(0.0082)***	(0.0082)***	(0.0082)***	(0.0082)***
log (Interpurchase Time)	0.0044	0.0045	0.0045	0.0043	0.0043
	(0.0025)*	(0.0025)*	(0.0025)*	(0.0025)*	(0.0025)*
-log(likelihood)	3059.2	3052.5	3056.6	3057.5	3062.4
AIC[2]	6142.4	6121.0	6129.1	6134.9	6148.8
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 satisfaction ratings, estimated from Equation 1 and 2

[2] AIC for our proposed model (Column (1)) is reduced to 6127.6 when the insignificant variables are excluded from the equation.

Table 7: Parameter Estimates: Product Image Evaluations (Study 1)

	with Product Images (1)	Proposed Model (2)
Rating Incidence		
Intercept	-0.4223	-0.7481
	(0.0491)***	(0.0152)***
Disconfirmation	0.2699	0.0341
	(0.0523)***	(0.0260)
Performance Inconsistency	0.3150	-0.0110
	(0.1226)**	(0.0617)
Product Image Evaluations	-0.0049	
	(0.0471)	
Number of Transactions Since the Last Rating	-0.0797	-0.0652
	(0.0041)***	(0.0016)***
Satisfaction Ratings		
Intercept	2.6454	2.6962
	(0.1033)***	(0.0560)***
Disconfirmation	-0.7771	-0.5883
	(0.0861)***	(0.0524)***
Performance Inconsistency	-0.8615	-0.9040
	(0.2089)***	(0.1263)***
Product Image Evaluations	-0.1039	
	(0.0836)	
Error Correlation (ρ_{12})	0.4781	0.4111
	(0.1251)***	(0.0635)***
Interpurchase Time		
Intercept	1.9985	1.5992
	(0.2015)***	(0.0979)***
Predicted Satisfaction	-0.0435	-0.0350
	(0.0243)*	(0.0106)***
Disconfirmation	0.1865	0.0549
	(0.0831)**	(0.0476)
Performance Inconsistency	-0.5623	0.0361
	(0.1971)***	(0.1133)
Product Image Evaluations	0.1270	
	(0.0783)	
Smoothing Factor (δ)	0.40	0.40
-log(likelihood)	5708.1	18535.7
AIC	11442.3	37095.4
Number of Observations	24723	141301

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

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

Table 8: Parameter Estimates: Across-Individuals Selection (Study 1)

	Across-Individuals (1)	Within-Individual (2)
Rating Incidence		
Intercept	-1.3840	-0.7481
	(0.0239)***	(0.0152)***
Disconfirmation	0.0423	0.0341
	(0.0387)	(0.0260)
Performance Inconsistency	0.0774	-0.0110
	(0.0834)	(0.0617)
Number of Transactions Since the Last Rating	-0.1187	-0.0652
	(0.0045)***	(0.0016)***
Satisfaction Ratings		
Intercept	2.3240	2.6962
	(0.1874)***	(0.0560)***
Disconfirmation	-0.3760	-0.5883
	(0.1122)***	(0.0524)***
Performance Inconsistency	-0.7580	-0.9040
	(0.2514)***	(0.1263)***
Error Correlation (ρ_{12})	0.1350	0.4111
	(0.0929)	(0.0635)***
Interpurchase Time		
Intercept	2.4183	1.5992
	(0.2156)***	(0.0979)***
Predicted Satisfaction	-0.0376	-0.0350
	(0.0142)***	(0.0106)***
Disconfirmation	0.1529	0.0549
	(0.0994)	(0.0476)
Performance Inconsistency	-0.2818	0.0361
	(0.2167)	(0.1133)
Smoothing Factor (δ)	0.40	0.40
Adjusted R-squared	0.0446	
-log(likelihood)		18535.7
AIC		37095.4
Number of Observations	141301	141301

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

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

Table 9: Parameter Estimates: Alternative Measures of Customer Expectation (Study 1)

	Most Recent (1)	Simple Average (2)	Proposed Model (3)
Rating Incidence			
Intercept	-0.7479	-0.7481	-0.7481
	(0.0152)***	(0.0152)***	(0.0152)***
Disconfirmation	0.0339	0.0247	0.0341
	(0.0216)	(0.0265)	(0.0260)
Performance Inconsistency	-0.0131	-0.0113	-0.0110
	(0.0617)	(0.0617)	(0.0617)
Number of Transactions	-0.0652	-0.0652	-0.0652
Since the Last Rating	(0.0016)***	(0.0016)***	(0.0016)***
Satisfaction Ratings			
Intercept	2.6974	2.7009	2.6962
	(0.0552)***	(0.0554)***	(0.0560)***
Disconfirmation	-0.3932	-0.6033	-0.5883
	(0.0441)***	(0.0530)***	(0.0524)***
Performance Inconsistency	-0.9035	-0.9113	-0.9040
	(0.1257)***	(0.1261)***	(0.1263)***
Error Correlation (ρ_{12})	0.4182	0.4166	0.4111
	(0.0635)***	(0.0636)***	(0.0635)***
Interpurchase Time			
Intercept	1.5999	1.5997	1.5992
	(0.0979)***	(0.0979)***	(0.0979)***
Predicted Satisfaction	-0.0344	-0.0348	-0.0350
	(0.0106)***	(0.0106)***	(0.0106)***
Disconfirmation	0.0953	0.0772	0.0549
	(0.0400)**	(0.0484)	(0.0476)
Performance Inconsistency	0.0294	0.0377	0.0361
	(0.1133)	(0.1133)	(0.1133)
Smoothing Factor (δ)	0.40	0.40	0.40
-log(likelihood)	18533.9	18535.2	18535.7
AIC	37091.8	37094.3	37095.4
Number of Observations	141301	141301	141301

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

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

Table 10: Descriptive Statistics: All Raters vs. Raters who Purchased 3 or More Times (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

(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

Table 11: Descriptive Statistics: Transactions with vs. without Satisfaction Ratings (Study 2)

(a) Transactions with satisfaction ratings (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

(b) Transactions without satisfaction ratings (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

Table 12: Parameter Estimates: Customer Rating and Repurchase Behavior (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 Ratings	
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 13: Consistent Results across Both Settings

(a) Quick Service Restaurant Industry

	Rating Incidence	Satisfaction Rating	Repurchase
Disconfirmation		Negative	
Performance Inconsistency		Negative	
Customer Satisfaction			Negative

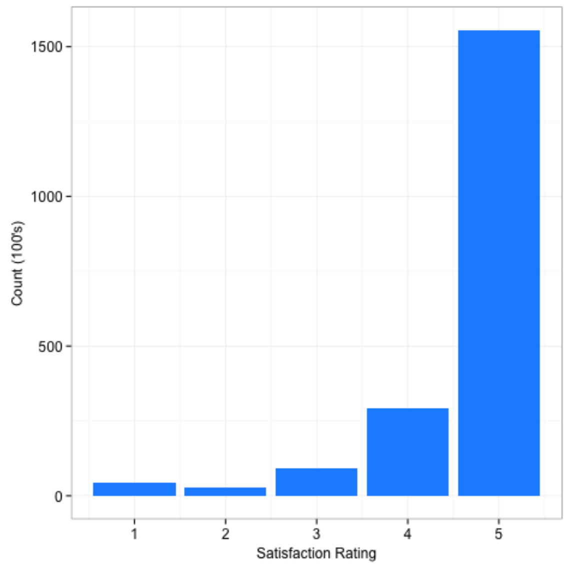
(b) Auto Rental Industry

	Rating Incidence	Satisfaction Rating	Repurchase
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

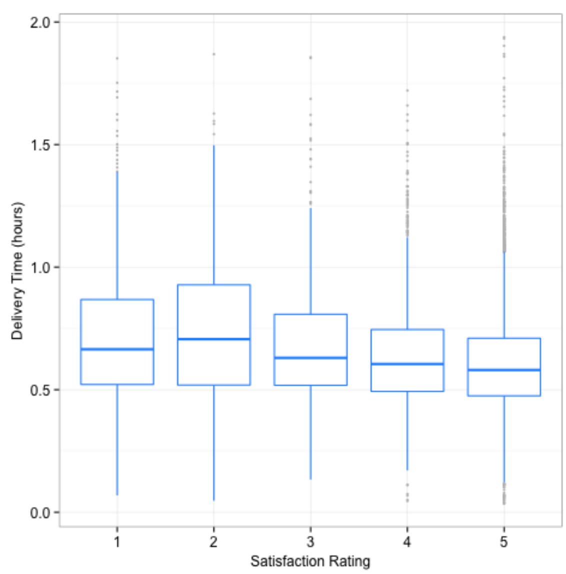
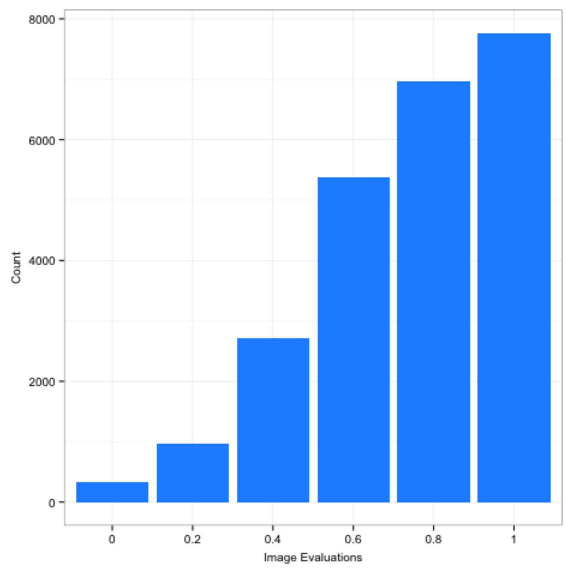


Figure 2: Average Per-day Product Image Evaluations (Study 1)

(a) Distribution of Average Per-day Product Image Evaluations



(b) Distribution of Interpurchase Time across Different Average Evaluations

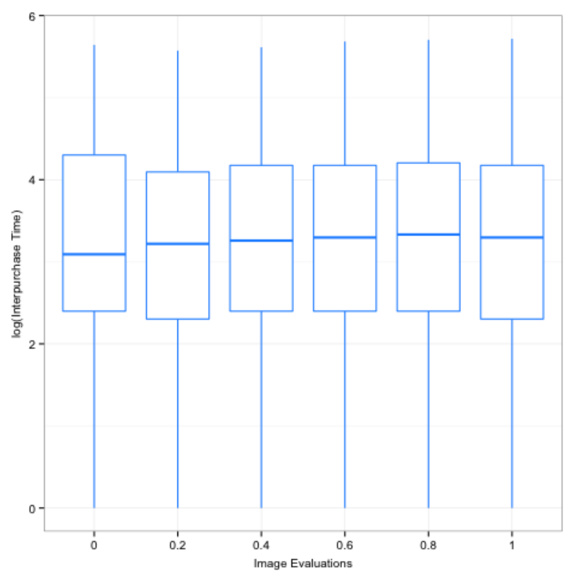
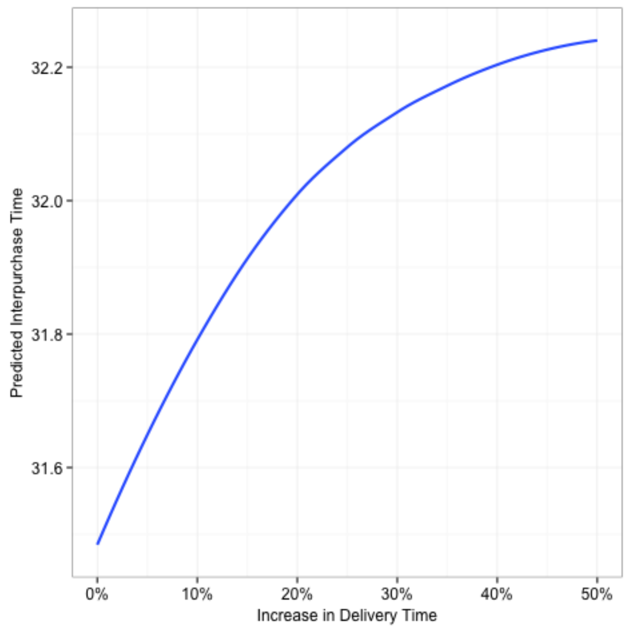


Figure 3: The Effect of Delay in Service (Study 1)

(a) on Interpurchase Time



(b) on Customer Satisfaction Ratings

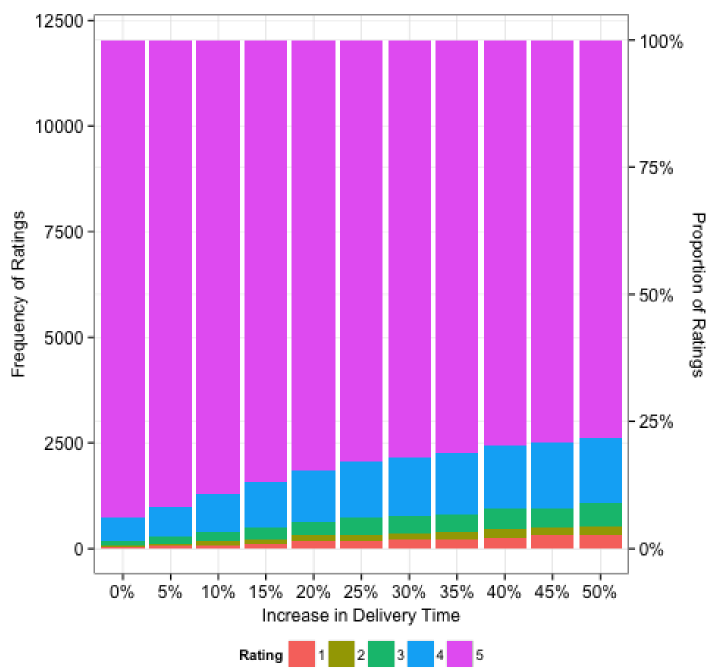


Figure 4: Distribution of Satisfaction Ratings and Free Upgrades (Study 2)