THE CAUSAL EFFECT OF SERVICE SATISFACTION ON CUSTOMER LOYALTY

By

Guofang Huang and K. Sudhir

May 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2177

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

The Causal Effect of Service Satisfaction on Customer Loyalty

Guofang Huang
Krannert School of Management, Purdue University, huan1259@purdue.edu

K. Sudhir
Yale School of Management, k.sudhir@yale.edu

May, 2019

We propose an instrumental-variable (IV) approach to estimate the causal effect of service satisfaction on customer loyalty, by exploiting a common source of randomness in the assignment of service employees to customers in service queues. Our approach can be applied at no incremental cost by using routine repeated cross-sectional customer survey data collected by firms. The IV approach addresses multiple sources of biases that pose challenges in estimating the causal effect using cross-sectional data: (i) the upward bias from common-method variance due to the joint measurement of service satisfaction and loyalty intent in surveys; (ii) the attenuation bias caused by measurement errors in service satisfaction; and (iii) the omitted-variable bias that may be in either direction. In contrast to the common concern about the upward common-method bias in the estimates using cross-sectional survey data, we find that ordinary-least-squares (OLS) substantially underestimates the casual effect, suggesting that the downward bias due to measurement errors and/or omitted variables is dominant. The underestimation is even more significant with a behavioral measure of loyalty–where there is no common methods bias. This downward bias leads to significant underestimation of the positive profit impact from improving service satisfaction and can lead to under-investment by firms in service satisfaction. Finally, we find that the causal effect of service satisfaction on loyalty is greater for more difficult types of services.

Key words: service satisfaction, customer loyalty, common-method bias, measurement error, cross-sectional data

---

1. Introduction

Service encounters are often referred to as "moments of truth"— instances that give the customer an opportunity to either form or change their impression about the firm. Service satisfaction is considered a forward-looking metric of the health of a firm's customer base because of its impact on outcomes like word of mouth, cross-selling and retention (e.g., Menezes and Serbin 1991, Anderson et al. 2004, Cronin and Taylor 1992, Parasuraman et al. 1985, Zahorik and Rust 1992). For this reason, firms routinely conduct surveys of customers to obtain their evaluations of service encounters (Zeithaml et al. 1996). To be sure, not only do pure service firms/organizations such as banks, hotels, restaurants and health-care providers conduct surveys to track their service performance, firms selling products also use such surveys to track performance on auxiliary services such as delivery, installation and customer support.

Despite the voluminous literature estimating the relationship between cross-sectional survey-based metrics of service satisfaction and customer loyalty (see extensive reviews in Shankar et al. 2003, Kumar et al. 2013), an enduring debate about whether increasing service satisfaction leads to greater retention and better financial results has continued.[1] One reason for this debate is that the estimated relationships between service satisfaction and loyalty are potentially biased due to multiple sources of potential bias (e.g., the common-method problem, errors in satisfaction measurement and omitted variables), whose aggregate impact is unknown a priori.

The goal of this paper is to propose an instrumental-variable (IV) approach for estimating the unbiased causal relationship between customer satisfaction in service encounters and loyalty. The IV approach exploits a common source of randomness, the availability of individual service employees, in the assignment of service employees to customers in service queues. Because the availability of individual service employees (of a certain qualification) at the time of a service request is independent of the waiting customer, we propose using the skill level of the assigned service employee as an instrument for service satisfaction to obtain the unbiased estimate of the causal relationship. The IV approach exploits common cross-sectional surveys of customers and can also be applied to estimate the causal relationship between service satisfaction and other customer outcome metrics, and thus is of high practical value at little incremental cost for firms.

Though the relationship between service satisfaction and the various metrics of loyalty is generally expected to be positive, the magnitude of the impact of service satisfaction on loyalty can vary significantly across different settings (Shankar et al. 2003). The variation could be across different types of service activities within a firm (e.g., online/offline; installation/delivery; check-in/room service), across firms within an industry (e.g, due to market share; brand strength and differentiation) and across industries (e.g., extent of competition). Given this variability in the magnitude of the relationship, managers increasingly

[1] Service satisfaction (quality) is typically conceptualized as the gap between "perceived quality" and "expected quality." The literature routinely uses service satisfaction and quality interchangeably because service quality can only be measured in terms of customer's satisfaction with the service encounter.

require *firm* and *context-specific* evidence of the financial soundness of investments in service satisfaction through its effects on customer loyalty and profitability (Zeithaml et al. 1996). Thus, our IV approach that can be applied within a firm is of significant value to managers seeking to determine the appropriate levels of investment in improving service within their respective firms.

For our empirical application, we consider two commonly used measures of loyalty—one based on surveys and another based on behaviors. The first metric is "willingness to recommend to a friend" (hereafter referred to by its common acronym, RTF)—a commonly used survey-based measure of loyalty used by many firms, where RTF is measured on a $1-10$ scale and higher numbers indicate greater likelihood of recommending to a friend.[2] The second metric "(the lack of) attrition" is a behavioral metric of loyalty. In our empirical application using data from the call center of a large credit-card issuer, we define attrition (the opposite of retention) as a customer canceling his/her card issued by the company. Our IV approach works with both the survey and behavioral measures of loyalty.

We now elaborate on how estimates of the relationship between service satisfaction and loyalty are typically affected by various sources of bias noted above.

1. Common methods bias

It is well-known that when multiple constructs are measured through self-reports of perceptions and impressions within the same survey (as is typically the case with measurements of customer satisfaction and RTF), one can have spurious correlations between these constructs due to response styles, social desirability and priming effects which are independent from the true casual relations among the constructs being measured. This bias known as "common methods bias" (Podsakoff et al. 2003, Kamakura 2010) can lead to substantial overestimation of the relationship between satisfaction and self-reported measures of loyalty such as RTF.

2. Attenuation bias due to measurement error

Satisfaction is a perception measure and can be measured only by surveying the customer who received the service. For the same true satisfaction level, the reported satisfaction

[2] RTF is the loyalty question underlying the "Net Promoter Score" (NPS) metric that is recommended as a predictor of future growth (Reichheld 2003, Reichheld and Covey 2006), with NPS = "% of customers with RTF $>= 9''$ – "% of customers with RTF$<=6$". Heskett and Sasser (2010) note that the RTF question is widely used in industry as a measure of loyalty due to its simplicity and intuitive appeal. In a Bloomberg report, Kaplan (2016) notes that over two-thirds of Fortune 1000 companies use the RTF question.

levels can vary across respondents due to, for example, inattention and differences in customers' response styles (Mittal and Kamakura 2001, Büschken et al. 2013).[3] The literature dealt with the measurement error problem by controlling for the moderating effects of customer characteristics (Mittal and Kamakura 2001). Though it is well-known that classical measurement errors lead to attenuation biases (i.e., the magnitudes of the effects being under-estimated) in the OLS estimates, there has been little acknowledgment in the customer satisfaction literature that the relationship between satisfaction and loyalty can be systematically underestimated due to measurement error in the survey measures of customer satisfaction.

3. Omitted Variables

More generally, there are likely some omitted variables that are correlated with both satisfaction and customer loyalty. For example, customers' (unobserved) expectations of service quality can affect both their satisfaction and their loyalty; and the unobserved triggers of service calls can also affect both customer satisfaction and loyalty. The sign of the bias caused by an omitted variable is specific to the omitted variable.

These sources of bias also make it challenging to even just determine the direction of the biases in the estimated effects of service satisfaction on loyalty from standard OLS regressions. In estimating the relationship between service satisfaction and stated loyalty (e.g., intent to repurchase, RTF), common methods bias and attenuation bias will both be present. While common methods bias leads to upward bias, measurement error leads to downward bias. Furthermore, in estimating the relationship between service satisfaction and the two metrics of loyalty, the existence of possibly multiple omitted variables further adds to the challenge. The omitted variables may cause biases in the estimates in either direction and the magnitude of the biases are typically unknown. Hence it is not even feasible to "sign" the direction of biases a priori.

We apply our IV approach to estimate the impact of service satisfaction using data of service-encounter surveys and internal records from a large credit-card issuer. We focus on answering the following key research questions in our analysis.

1. What are the causal effects of service satisfaction on RTF (stated loyalty) and attrition (behavioral loyalty)?

[3] Note that the measurement errors in the dependent variables do not cause any biases in the estimates.

2. Do the causal effects of service satisfaction on customer loyalty vary with the difficulty and/or the importance of the service requests?
3. Does obtaining the unbiased estimates using the IV approach have a significant impact on managerial actions such as investments in service satisfaction and customer targeting with premier service?

Our results show that the IV estimates of the causal impact on customer loyalty are significantly larger than the counterparts obtained through standard OLS regressions. For behavioral loyalty (attrition), the IV estimates are around twice as large in magnitude as the corresponding OLS estimates. The difference between the IV and OLS estimates are managerially significant. Our estimates suggest that a 0.4-point increase in satisfaction for a single service call can lower the probability of losing the calling customer in the following 18 months by around 0.9 percentage point (ppt) on average, which implies an increase in the profit per customer by \$15.1 (\$7.7) according to the IV (OLS) estimates.[4] Thus, basing decisions on the OLS estimates, the company would significantly under-invest in service quality.

Our IV estimates also show that the causal impact of service satisfaction is larger for calls that are more difficult to handle or more important to customers. This differential impact of customer satisfaction suggests that the company may consider creating elite teams of reps and/or provide stronger incentives to improve the service quality of these more challenging/important types of calls.

The rest of the paper is organized as follows. Section 2 describes our data. Section 3 describes our identification strategy and estimation results. Section 4 presents additional results from applying our IV approach to studying the heterogeneity in the causal impact of satisfaction by call types. Section 5 discusses the managerial implications and Section 6 concludes.

2. Data

We begin with a description of our data. The first part of our data consists of all the responses to the standard service satisfaction surveys conducted by a large credit card issuer on customers who called and spoke to a service representative (hereafter "rep") at

[4] A 0.4-point increase in average service satisfaction can result from increasing the call-handling rep's skill level from that of a 5th percentile rep to that of a 95th percentile rep. The numbers are calculated under the assumption that the length of the remaining customer relationship is *at most* 10 years for all customers.

its customer service center from March 2008 to December 2009. The survey asks customers about their satisfaction with the service of the call center rep that they interacted with and their likelihood of recommending the company's card products to their friends (the "RTF" score). The survey data also include the identity of the rep that handled each call and the reason for each call. We use the survey data to construct proxies for the skill levels of reps.

The focus of our empirical analysis is on the roughly 42,000 customers who called in January 2009 and responded to the surveys after their calls. For these customers (but not those who called and responded to surveys in other months of the around two-year period), our data also include the internal descriptive and behavioral data provided by the credit card issuer. For each rep appeared in the data of January 2009, we compute two average satisfaction scores, one using the survey data from from March–November 2008 and the other using the survey data from April–December 2009. We also compute each rep's average satisfaction score separately for each type/reason using the survey data from March–November 2008 to measure the rep skill level for each type of calls. These measures are our proxies for each rep's skill level.

Table 1 reports the summary statistics of our data on customers who made called in January 2009. Satisfaction is measured on a scale of one to five. Overall, customers are quite satisfied, with an average of 4.28. RTF is measured on a scale of one to ten, with higher numbers indicating higher likelihood of recommendation to a friend, and 10 indicate "will definitely recommend to a friend." The average of RTF across all calls is 8.54. Reflecting the company's position as a premier credit card issuer, the average size of wallet is large, and the company has a very high average share of wallet at 55%.[5] The FICO score is very high and the average age of the card holder is also high at about 57 years. Attrition rate—the percentage of the cards in our data being canceled upon customers' request or by the company for being inactive—over the 18 months starting from Feb 2009 is 9%.[6]

We augment the data of January 2009 with the skill level (as measured by our proxies) of the corresponding rep for each call. The last three rows of Table 1 reports the summary

[5] The size of wallet is defined as the total spent by a customer on all credit/debit cards over a year. The share of wallet is defined as the total spent on the company's cards by a customer divided by the customer's size of wallet.

[6] We cannot identify those who cancel because they switched to another card issued by the company. We learned from the company that such cases should be only a very small share of the attrition. Furthermore, service satisfaction most likely does not have a significant impact on a customer's decision on whether to continue to use the card in our data or switch to a different card issued by the company. Thus, the impact of the measurement issue on our estimates should be very limited.

statistics of the call-level rep skill for the customer data of January 2009. The summary statistics show significant variations in the rep skill level across calls. The standard deviation is greater for call-type specific rep skill level, reflecting additional heterogeneity in rep skill across call types.

Table 1 Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Satisfaction	4.28	1.02	1	5	42337
RTF	8.54	2.34	1	10	42337
Customer Tenure (years)	10.95	10.47	0	51	42338
Size of Wallet ($1,000)	32.78	62.34	0	6285	42338
Share of Wallet (%)	54.49	34.86	0	100	42143
FICO Score	757.56	61.64	423	997	41948
Female	0.24	0.43	0	1	42338
Male	0.28	0.45	0	1	42338
Age	56.61	14.79	19	117	37099
Customer Attrition within 18 months	0.09	0.29	0	1	42338
Rep Avg. Sat (before).	4.31	0.25	2.42	5	41965
Rep Avg. Sat (after)	4.25	0.31	1.83	5	38093
Rep-Call Type Avg. Sat. (before)	4.32	0.65	1	5	34357

Note: 1) The unit of observation is a call with survey result in our sample of Jan 2009; 2) Rep Avg. Sat (Rep-Call Type Avg. Sat. (before)) is the average satisfaction rating (by call type) of the rep handling a call; 3) the means of Female and Male do not sum up to one because the gender information is missing for some customers.

There is significant variation in both service satisfaction and the outcome metrics across call types. Table 2 shows the means of satisfaction, RTF, attrition rate (in the following 18 months) and the share of each call type. Most saliently, the first four call types (in boldface) are the ones where reps may have to say no to customer requests; not surprisingly, both the average satisfaction and RTF tend to be much lower for these calls, than for other call types where the service is mostly assessed by the quality of the experience and reps are mostly able to satisfy customer requests.

Figures 1a and 1b shows that the distributions of satisfaction ratings and RTF by call types have similar patterns as we noted in Table 2. In particular, the figures show that there are significant differences in the distributions across call types. Calls about APR and line of credit tend to have significantly lower ratings for both satisfaction and RTF. It is, thus, important to control for the fixed effects of call types in estimating the causal effects of service satisfaction on the outcome metrics.

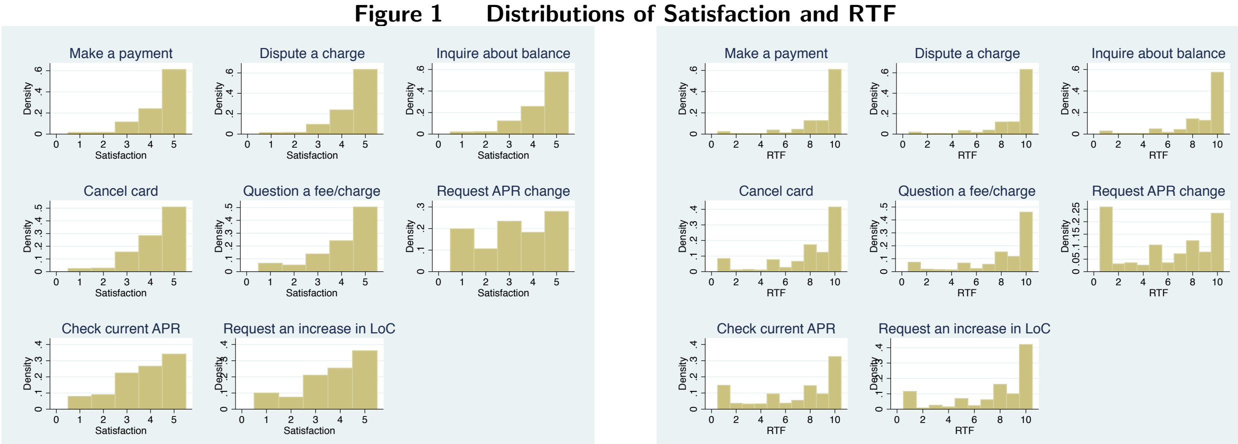
Many other unobserved factors, such as customers' satisfaction with other product features, can also affect both service satisfaction and the outcome metrics. For example, figure 2 shows that the distributions of satisfaction ratings and RTF for calls to request changes

Table 2 Mean Survey Outcomes and Attrition Rate by Call Types

Call types	Satisfaction	RTF	Attrition	% of Calls
Request a change in APR	3.239	5.801	0.094	2.75
Check current APR	3.684	6.745	0.089	2.3
Request an increase in line of credit	3.695	7.567	0.067	0.92
Check available line of credit	3.900	7.654	0.072	0.79
Make a payment or make issuer aware of a payment	4.418	8.905	0.061	13.32
Dispute an inappropriate or incorrect charge	4.460	8.983	0.060	12.16
Inquire about balance/account/bill	4.348	8.758	0.075	11.5
Question a fee or charge	4.071	7.984	0.093	6.76
Other reason	4.143	8.401	0.085	4.44
Clarify an unrecognized charge	4.540	9.101	0.055	3.99
Check recent charges/recent credits	4.474	9.023	0.059	3.9
Cancel Card	4.227	7.851	0.603	2.9
Card products or benefits	4.222	8.529	0.116	2.44
Request a copy of statement or a specific charge	4.292	8.720	0.061	2.2
Replace a lost, stolen, or damaged Card	4.518	9.026	0.069	1.89
Find out why a charge was denied	4.109	8.387	0.040	1.79
Inquire about user ID or password	4.394	8.722	0.056	1.74
Membership Rewards	4.139	8.424	0.105	1.68
Other rewards programs such as Delta Sky	4.001	7.885	0.160	1.51
Change or correct address/email/phone	4.454	8.895	0.073	1.43
Check on the status of a renewal/replacement card	4.347	8.671	0.060	1.43
Help locating information on the website	4.306	8.658	0.057	1.4
Charge refused	4.380	8.698	0.078	1.33
Fraud Issues (identity theft, stolen identity)	4.365	8.901	0.070	1.25
Traveling out of/back in town/country	4.586	9.015	0.048	1.21
Balance transfer	4.079	8.157	0.111	1.2
Change card products	4.273	8.493	0.163	1.12
Check payment due date	4.451	8.843	0.053	1

Note: Besides the two highlighted call reasons related to the line of credit, only call reasons that are at least 1% of call volume are listed.

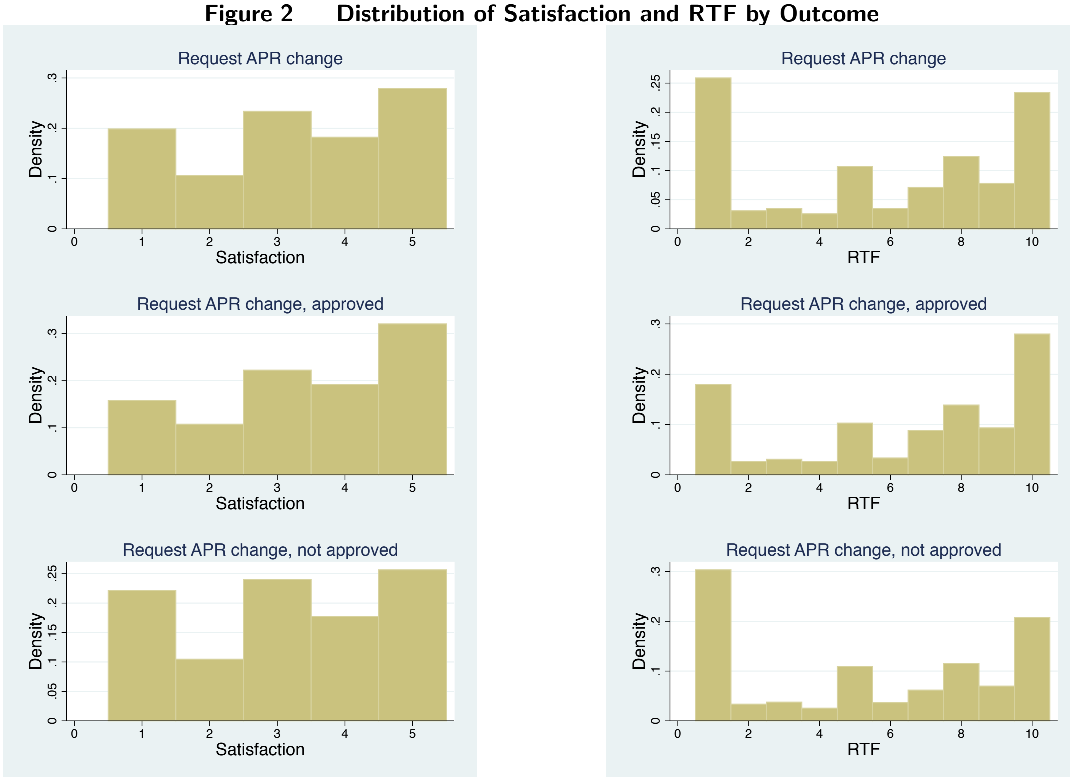
Figure 1 Distributions of Satisfaction and RTF



(a) Satisfaction by Call Reasons **(b) RTF by Call Reasons**

Note: The unit of observation is a call with survey result.

Figure 2 Distribution of Satisfaction and RTF by Outcome



(a) Satisfaction by Call Reasons

(b) RTF by Call Reasons

Note: The unit of observation is a call with survey result.

in APR also vary by whether reps responded positively or negatively to such requests. Here we define a request to change APR being "approved" if and only if there was a downward adjustment in the APR for a customer from January to February 2009.[7] As expected, when the customer request was not acceded both satisfaction and RTF were more negatively skewed, relative to when the APR reduction request was approved.

3. Empirical Analysis

We begin this section with a description of our empirical strategy. Next, we present our empirical results, comparing OLS estimates to those estimated with our IV strategy to demonstrate the bias in the OLS estimates and gain insights into the direction and size of the bias. Then, we discuss in detail how reps are assigned to calls and the appropriateness of our IV approach. Last, we discuss the robustness of our main findings.

[7] The APRs for the customers in our Jan 2009 samples are provided by the same credit card issuer mentioned above.

3.1. The Instrumental-Variable Approach

We focus the discussion of our IV strategy on the estimation of the following structural equation:

$$RTF_{it} = \alpha Sat_{it} + Z_{it}\beta + v_{it} \tag{1}$$

where customer i called customer service at time t; Sat_{it} is customer i's satisfaction with the service call; and Z_{it} is a vector of exogenous control variables and v_{it} is a scalar random variable. Let customer satisfaction be determined as follows:

$$Sat_{it} = h(s_{r(it)}) + Z_{it}\gamma + u_{it} \tag{2}$$

where s_r is the skill level of rep r; $r(it)$ indicates the rep who handled customer i's call at time t; $h()$ is an increasing function; and u_{it} is a scalar random variable.

The objective is to estimate α, the causal impact of customer satisfaction on RTF. As we noted, the survey metric of service satisfaction is a noisy measure of the customer's true level of satisfaction with the service encounter, and is potentially correlated with the error term (v_{it}) due to unobserved customer-specific expectations, response style and other omitted factors. A valid instrument here should be correlated with the customer's true satisfaction level, but uncorrelated with the error term in the regression.

We propose using the skill level of the assigned rep as an instrument for Sat_{it}: conditional on the profile (which the company uses to clarify a rep's qualification for handling calls regarding certain card products, customers and special issues) of the assigned rep, the assigned rep (and his/her skill level in particular) is independent of the calling customer (and, consequently, RTF_{it}). In practice, exogenous external measures of rep skill levels make the ideal instruments. Such exogenous measures may also be available from hiring tests or interview ratings of the rep, if sufficient correlation between these skill measures and overall satisfaction can be established.

The conditional independence property of rep skills and RTF is satisfied in our setting because the assignments of reps to calls are automated based on the reps' random availability. The conditional exogenous assignment of reps is confirmed by the company's managers with knowledge about the assignment process. Later in subsection 3.5, we discuss the rep assignment process in greater detail and provide empirical evidence for exogeneity in rep assignments.

Given that we do not have external measures of rep skill levels, we use rep-level average satisfaction ratings as proxies for reps' skill levels. To avoid the problem of certain contemporaneous factors affecting the service satisfaction with a rep in the same period, we use the average satisfaction rating of each rep in a past/future period as proxies for the skill level of reps. More specifically, let T indicate the period of the data that we use in estimating equation (1), and let T' and T'' indicate a past period and a future period (i.e., $\max T' < \min T$ and $\min T'' > \max T$), respectively.[8] Then our primary proxy for the skill level of rep $r(it)$ is: $\overline{Sat}_{r(it)b} = \frac{1}{N'_{r(it)}} \sum_{j \in C'_{r(it)}} Sat_{jt'}$, where $C'_{r(it)} = \{j | j \neq i, r(jt') = r(it), t' \in T'\}$ is the set of customers (in the survey data) whose calls were also answered by rep $r(it)$ in period T', $N'_{r(it)}$ is the number of survey observations available in period T'. Another proxy that we consider for rep skill is $\overline{Sat}_{r(it)a} = \frac{1}{N''_{r(it)}} \sum_{j \in C''_{r(it)}} Sat_j$, where $C''_{r(it)} = \{j | j \neq i, r(jt'') = r(it), t'' \in T''\}$ and $N''_{r(it)}$ is the number of survey observations available in period T''. We will refer to $\overline{Sat}_{r(it)b}$ and $\overline{Sat}_{r(it)a}$ as "Rep Avg. Sat (before)" and "Rep Avg. Sat (after)" later in the discussion of our empirical findings.

We begin with a discussion of issues to be considered in applying our IV approach with the above proxies for the rep's skill. We focus our discussion on using $\overline{Sat}_{r(it)b}$ as the IV; the same discussion applies to using $\overline{Sat}_{r(it)a}$ as the IV.

To use $\overline{Sat}_{r(it)b}$ as an IV for Sat_{it}, we require that $Cov(v_{it}, \overline{Sat}_{r(it)b}) = 0$.[9] The main issue that we need to attend to here is that the existence of group (card product/call type) fixed effects may make the unconditional covariance condition restrictive. With the group fixed effects, we have: $v_{it} = \phi_{g(i)} + \epsilon_{it}$, $u_{it} = \varphi_{g(i)} + \varepsilon_{it}$ for v_{it} and u_{it} in equations (1) and (2) respectively, where $\varphi_{g(i)}$ and $\varphi_{g(i)}$ are the fixed effects of group $g(i)$ (i.e. the group pertaining to customer i) in the two equations and ϵ_{it} and ε_{it} are idiosyncratic errors. We assume that $Cov(\epsilon_{it}, \varepsilon_{it'}) = Cov(\epsilon_{it}, Z_{jt'}) = 0$ for $i \neq j$ and $t \neq t'$, which seems reasonable because ϵ_{it} and $\varepsilon_{it'}$ concern different customers ($i \neq j$) at different points in time and the group fixed effects have been accounted for. By the definition of $\overline{Sat}_{r(it)b}$, we have:

$$\overline{Sat}_{r(it)b} = h(s_{r(it)}) + \frac{1}{N'_{r(it)}} \sum_{j \in C'_{r(it)}} Z_{jt'}\gamma + \frac{1}{N'_{r(it)}} \sum_{j \in C'_{r(it)}} (\varphi_{g(j)} + \varepsilon_{jt'}).$$

[8] Note that T, T' and T'' indicate sets of dates.

[9] The covariance condition follows if $Cov(RTF_{it}, Sat_{jt'} | Sat_{it}, Z_{it}) = 0$ for $j \neq i$, $r(jt') = r(it)$ and $t' \in T'$. By the definition of $\overline{Sat}_{r(it)b}$, we do not require $Cov(RTF_{it}, Sat_{it'} | Sat_{it}, Z_{it}) = 0$ for $t' < t$, which is similar to the assumption required by Gordon and Hartman (2012). With $t' < t$, $Cov(RTF_{it}, Sat_{it'} | Sat_{it}, Z_{it}) = 0$ may not hold because $Sat_{it'}$ can directly affect RTF_{it}, and thus $Sat_{it'}$ should not be included in calculating $\overline{Sat}_{r(it)b}$.

Let $\bar{\varphi}_{r(it)} \equiv \frac{1}{N'_{r(it)}} \sum_{j \in C'_{r(it)}} \varphi_{g(j)}$. Then, the identification condition $Cov(v_{it}, \overline{Sat}_{r(it)b}) = 0$ is equivalent to $Cov(\phi_{g(i)}, h(s_{r(it)}) + \bar{\varphi}_{r(it)}) = 0$, which can be restrictive in the presence of group fixed effects because i) $\phi_{g(i)}$ and $\varphi_{g(i)}$ can be correlated and $g(j) = g(i)$ (and thus $\varphi_{g(j)} = \varphi_{g(i)}$) for some $j \in C_{r(it)}$; and ii) the average skill level of the reps responsible for a group of customers may be correlated with certain unobserved characteristics of the group.

To deal with the above issue, we control for the group fixed effects. Then, we have $Cov(\phi_{g(i)}, h(s_{r(it)}) + \bar{\varphi}_{r(it)} | g(i)) = 0$. Thus, we have $Cov(v_{it}, \overline{Sat}_{r(it)b} | g(i)) = 0$ and $\overline{Sat}_{r(it)b}$ is a valid IV for Sat_{it} once we control for the group fixed effects.

In cases where one might be concerned that $Cov(\epsilon_{it}, \varepsilon_{jt'}) \neq 0$, for $t - \Delta t < t' < t$ (but $Cov(\epsilon_{it}, \varepsilon_{jt'}) = 0$, for $t' < t - \Delta t$) for some $\Delta t > 0$, we can choose T' such that $\max T' < \min T - \Delta t$ for some Δt, when using $\overline{Sat}_{r(it)b}$ as an IV. Note that this serial-correlation issue would not be relevant if we had independent external measurements of rep skills.

In our empirical application, the primary IV we focus on is “Rep Avg. Sat (before)”, the rep average customer satisfaction calculated using the survey data from March-November 2008. We do not use surveys from December 2008 in calculating “Rep Avg. Sat (before)” to guard against the potential serial correlation in the error terms mentioned above. We calculate “Rep Avg Sat (after)” using the survey data from April-December 2009. To capture the differences in service quality of the same rep across different types of service requests, we also calculate “Rep-Call Type Avg Sat (before)”, the average rep satisfaction for each particular type of call using the surveys from March-November 2008. The first two IVs are overall measures of each rep’s skill level, whereas the third IV measures each rep’s skill for handling each type of call. The correlation between the first and third rep skill measures is 0.39. We note that even though the Rep-Call Type Avg Sat may more directly affect customer satisfaction, it is also measured with lower precision due to the smaller sample size as there are much fewer calls of each type. The additional IVs allow us to conduct over-identification tests to test the exogeneity of the proposed IVs.

We report our empirical results in the next three subsections, starting with first-stage regressions and then the second-stage ones. We control for the fixed effects of card product by call type in all regressions.

3.2. Factors Determining Service Satisfaction

We report the first-stage regression results on the factors determining service satisfaction in Table 3. These results shed light on the relative importance of the various factors in

Table 3 Satisfaction with Service and Rep Skill

	Dependent Variable: Satisfaction					
	(1)	(2)	(3)	(4)	(5)	(6)
Rep Avg. Sat (before)	0.573***			0.570***		
	(0.0450)			(0.0460)		
Rep Avg. Sat (after)		0.524***			0.520***	
		(0.0301)			(0.0311)	
Rep-Call Type Avg. Sat (before)			0.122***			0.121***
			(0.0146)			(0.0146)
Customer Tenure (years)				-0.000296	0.0000467	-0.000149
				(0.000600)	(0.000607)	(0.000688)
Size of Wallet ($1,000)				0.0000974	0.000109	0.0000832
				(0.000101)	(0.000104)	(0.000110)
Share of Wallet (%)				0.000536***	0.000394*	0.000651***
				(0.000147)	(0.000157)	(0.000174)
FICO Score				0.000427***	0.000470***	0.000445***
				(0.000105)	(0.000116)	(0.000110)
Constant	1.810***	2.053***	3.772***	1.468***	1.689***	3.402***
	(0.194)	(0.128)	(0.0629)	(0.193)	(0.144)	(0.104)
Observations	40810	38093	32846	40307	37624	32476
R^2	0.0207	0.0264	0.0063	0.0221	0.0277	0.0078
(Incremental) F statistics	162.2	303.1	70.2	153.6	280.2	69.1

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

determining service satisfaction. They also show the power of the IVs that we proposed above. The results in columns (1)-(3) of Table 3 show that the rep's skill is an important factor in determining service satisfaction. The estimates in column (1) shows that the reported service satisfaction increases by 0.57 on average when Rep Avg. Sat (before) increases by one, and Rep Avg. Sat (before) alone explains 2.1% of the variations in service satisfaction. The results from column (2) are very similar to those from column (1), showing that Rep Avg. Sat (before) and Rep Avg. Sat (after) work similarly well as proxies for rep skill. Meanwhile, in contrast to the other two proxies, Rep-Call Type Avg. Sat (before) explains only about 0.6% of the variations in Satisfaction, suggesting it may work less well as an instrument for Satisfaction. Across all specifications, the causal impacts of rep skill, as measured by the three proxies, on Satisfaction are statistically significant at the 0.1% level.

In contrast, the main customer characteristics, such as customer tenure and share of wallet, explain much less (around 0.16%) of the variations in service satisfaction as seen in columns (4)-(6). Customers with higher shares of wallet and FICO scores reported higher satisfaction with their service experience. Table A1 in the Appendix shows that female customers also rate their satisfaction somewhat higher on average, but age does not have a significant relationship with satisfaction ratings. The relationships between customer characteristics and the reported service satisfaction can be due to, for example, the heterogeneity in customers' response styles or actual preferences.

3.3. The Causal Effect of Satisfaction on RTF

The regressions in Tables 4a show the estimated relationship between Satisfaction and RTF. Columns (1) and (4) present the OLS estimates; while columns (3) and (6) present the corresponding IV (2SLS) estimates with Rep Avg. Sat. (before) as the IV for Satisfaction. The IV estimates show a significant positive *casual* effect of satisfaction on RTF: a one point increase in satisfaction leads to a 1.9-point increase in RTF. Controlling for the customer characteristics in column 6 causes little change in the estimated effect of satisfaction. The relatively large R^2, 0.191, in column 3 shows that service satisfaction is a major factor in determining RTF.

Columns (2) and (5) in Table 4a are reduced-form OLS regressions that include Rep Avg. Sat. (before) directly in the regressions. The estimates show there is a significant positive causal effect of assigning a more skillful rep on RTF. Estimates in column (2) show that the rep skill (and, thus customer satisfaction) in a single service encounter can explain at least 1.5% of the variations in RTF. This estimate is economically significant in terms of its magnitude given the relatively low cost to the firm for a single call.

The comparison of OLS and IV estimates in Table 4a shows that, in spite of the significant explanatory power of satisfaction, OLS significantly *underestimates* the impact of Satisfaction on RTF. The finding is a bit surprising, because researchers typically are more concerned about the upward bias caused by the common-method problem.

The instruments we propose are not "weak instruments." For the first stage of the 2SLS regressions reported in columns (3) and (6) in Table 4a, the partial R^2 of Rep Avg. Sat (before) is 0.021 and 0.020 respectively, and the corresponding F statistics (for testing the null hypothesis of the first-stage coefficient of Rep Avg. Sat (before) being zero) are 162 and 154 respectively. The large F statistics ensure that Rep Avg. Sat (before) is not a weak instrument for Satisfaction based on the F-statistic test in Staiger and Stock (1997). Furthermore, satisfaction is indeed endogenous in the RTF equations, as we have conjectured above. The robust regression-based test of exogeneity suggested by Wooldridge (1995) rejects the null hypothesis of Satisfaction being exogenous in the RTF equations at the 0.1% level. The corresponding F statistics are 40 and 35.7 for the regressions reported in columns three and six of Table 4a, respectively.

Table 4a Customer Satisfaction and RTF

	Dependent Variable: RTF					
	OLS	OLS	IV	OLS	OLS	IV
Satisfaction	1.204***		1.910***	1.199***		1.902***
	(0.0307)		(0.100)	(0.0313)		(0.105)
Rep Avg. Sat (before)		1.094***			1.085***	
		(0.112)			(0.117)	
Customer Tenure (years)				0.00439***	0.00421**	0.00477***
				(0.000995)	(0.00129)	(0.000997)
Size of Wallet ($1,000)				-0.000117	-0.0000743	-0.000260
				(0.000157)	(0.000237)	(0.000167)
Share of Wallet (%)				0.00251***	0.00310***	0.00208***
				(0.000369)	(0.000424)	(0.000389)
FICO Score				-0.0000104	0.000530*	-0.000281
				(0.000211)	(0.000257)	(0.000216)
Constant	3.385***	3.820***	0.363	3.235***	3.248***	0.455
	(0.131)	(0.485)	(0.428)	(0.199)	(0.450)	(0.425)
Observations	42337	40810	40810	41814	40307	40307
R^2	0.2817	0.0146	0.1907	0.2830	0.0178	0.1928
First stage partial R^2			0.021			0.020

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table 4b Customer Satisfaction and RTF, IV Estimates

	Dependent Variable: RTF					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfaction	1.910***	1.821***	1.670***	1.902***	1.813***	1.648***
	(0.100)	(0.0865)	(0.168)	(0.105)	(0.0892)	(0.173)
Customer Tenure (years)				0.00477***	0.00415***	0.00506***
				(0.000997)	(0.000977)	(0.00108)
Size of Wallet ($1,000)				-0.000260	-0.000223	-0.000233
				(0.000167)	(0.000162)	(0.000177)
Share of Wallet (%)				0.00208***	0.00222***	0.00227***
				(0.000389)	(0.000394)	(0.000420)
FICO Score				-0.000281	-0.000369	-0.000121
				(0.000216)	(0.000218)	(0.000236)
Constant	0.363	0.741*	1.390	0.455	0.897*	1.408*
	(0.428)	(0.370)	(0.721)	(0.425)	(0.367)	(0.709)
Observations	40810	38093	32846	40307	37624	32476
R^2	0.191	0.210	0.232	0.193	0.211	0.237
IVs	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)	Rep-Call Type Avg. Sat. (before)	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)	Rep-Call Type Avg. Sat. (before)
First stage partial R^2	0.021	0.026	0.006	0.020	0.026	0.006

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table 4c Testing the Exogeneity of IVs by Using Over-identification Tests

	Dependent Variable: RTF					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfaction	1.731***	1.826***	1.398*	1.732***	1.812***	1.303
	(0.182)	(0.128)	(0.669)	(0.188)	(0.127)	(0.690)
Rep-Call Type Avg. Sat (before)	-0.0112	-0.0144		-0.0141	-0.0168	
	(0.0253)	(0.0232)		(0.0254)	(0.0233)	
Rep Avg. Sat (after)	0.0339		0.152	0.0286		0.181
	(0.0871)		(0.243)	(0.0880)		(0.251)
Rep Avg. Sat (before)		-0.0378	0.132		-0.0319	0.170
		(0.0978)	(0.299)		(0.0982)	(0.307)
Customer Tenure (years)				0.00448***	0.00448***	0.00451***
				(0.00105)	(0.00107)	(0.000995)
Size of Wallet ($1,000)				-0.000225	-0.000224	-0.000232
				(0.000176)	(0.000175)	(0.000192)
Share of Wallet (%)				0.00222***	0.00218***	0.00243***
				(0.000455)	(0.000452)	(0.000498)
FICO Score				-0.000215	-0.000248	-0.0000406
				(0.000250)	(0.000246)	(0.000347)
Constant	1.029*	0.941*	1.336	1.064*	1.016**	1.323*
	(0.452)	(0.368)	(0.696)	(0.439)	(0.387)	(0.559)
Observations	29810	29810	29810	29469	29469	29469
R^2	0.2183	0.1982	0.2718	0.2182	0.2015	0.2777
IVs	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)	Rep-Call Type Avg. Sat. (before)	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)	Rep-Call Type Avg. Sat. (before)

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table 4b reports the IV estimates of the RTF equation, using one of the three proposed IVs for satisfaction in each column. The point estimates of the coefficient of Satisfaction are similar for the three IVs, showing the robustness of our IV strategy. Meanwhile, the estimates using “Rep-Call Type Avg Sat (before)” as the IV is less accurate than those using the other two IVs, which is not surprising given the significantly smaller partial R^2 of “Rep-Call Type Avg Sat (before)” in the first-stage regressions (Table 3). The results suggest that the rep average ratings are actually less noisy measurement of each rep’s relevant skill—there are far fewer survey observations per rep for each specific call type and each rep’s skills for different call types are highly correlated.

Over-identification tests (Wooldridge 1995) cannot reject the exogeneity of the IVs at any standard significance level. Table 4c shows the exogeneity tests for the IVs by directly including the additional IVs in the second stage regressions. None of the IVs included in the second stage regressions are close to being statistically significant. These results confirm the exogeneity of the proposed IVs and show the validity of our proposed IV approach for obtaining consistent estimates of the causal relationship between service satisfaction and RTF (the survey-based metric of loyalty).

3.4. The Causal Effect of Satisfaction on Customer Loyalty

The results in this subsection show that OLS significantly underestimates the impact of customer satisfaction on loyalty—as measured by retention or lack of attrition. Table 5a presents our estimates of the impact of satisfaction on attrition. Wooldridge's (1995) robust endogeneity test shows that satisfaction is also endogenous in the attrition equation. The estimates show: 1) satisfaction has a significant negative causal impact on attrition; 2) OLS significantly underestimates the magnitude of the impact. The OLS estimates of satisfaction's impact significantly underestimate the true effect and are only about *one half* or less of the corresponding IV estimates. The IV estimates in column 6 show that a one point increase in Satisfaction leads to a decrease of 2.3 ppts in the attrition rate in the following 18 months. Given this being the result of a single encounter with customer service, the effect is quite large.

Table 5a Customer Satisfaction and Attrition in the Following Eighteen Months

	Dependent Variable: Attrition					
	OLS	OLS	IV	OLS	OLS	IV
Satisfaction	-0.0128***		-0.0301*	-0.0118***		-0.0232*
	(0.00152)		(0.0124)	(0.00150)		(0.0114)
Rep Avg. Sat (before)		-0.0172*			-0.0132*	
		(0.00723)			(0.00673)	
Customer Tenure (years)				-0.00280***	-0.00282***	-0.00282***
				(0.000326)	(0.000337)	(0.000335)
Size of Wallet ($1,000)				-0.000135**	-0.000134**	-0.000132**
				(0.0000450)	(0.0000456)	(0.0000450)
Share of Wallet (%)				-0.000795***	-0.000815***	-0.000802***
				(0.0000706)	(0.0000726)	(0.0000734)
FICO Score				0.000253***	0.000255***	0.000265***
				(0.0000426)	(0.0000431)	(0.0000427)
Constant	0.146***	0.166***	0.221***	0.0258	0.0329	0.0670
	(0.00649)	(0.0312)	(0.0530)	(0.0277)	(0.0437)	(0.0559)
Observations	42337	40810	40810	41814	40307	40307
R^2	0.0022	0.0002	.	0.0227	0.0214	0.0214
First stage partial R^2			0.021			0.020

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions; and 4) we use the linear probability model here for its flexibility to control a rich set of fixed effects.

Service satisfaction explains much less of the variation in attrition (as the prior literature anticipated) than for RTF (0.2% vs. 28% in column one of Tables 5a and 4a respectively). Rep skill also explains much less variation in attrition than that in RTF (0.02% vs. 1.5%). The gap may be driven by both the difference between attrition and RTF and that between behavior and intent.

Table 5b shows estimates of the attrition equation using the three different IVs. The point estimates of the satisfaction coefficient are reasonably similar, and the differences among them are statistically insignificant. The lack of statistical significance of the point estimates in columns (3) and (6) is likely due to the fact that "Rep-Call Type Avg Sat" is a relatively weaker IV (as shown by the smaller first stage partial R^2 of the IV). The consistent results from using different IVs show again the robustness of the proposed IV approach.

Given the findings above, we consider only Rep Avg. Sat (before) and Rep Avg. Sat (after) as the potential IVs for Satisfaction in our over-identification tests. We similarly implement the tests by including the extra IVs in the second-stage regressions (Wooldridge (1995)). The estimates reported in Table 5c show that we again cannot reject the exogeneity of the IVs at any standard levels. Including the two IVs in the tests leads to a smaller sample, because the two IVs are missing for two different sets of reps. The estimated coefficients of Satisfaction become insignificant in the test regressions, likely due to the smaller sample, the correlation between the two IVs and the more moderate impact of satisfaction on attrition (relative to on RTF).

Table 5b Customer Satisfaction and Attrition in the Following Eighteen Months, IV Estimates

	Dependent Variable: Attrition					
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfaction	-0.0301*	-0.0295**	-0.0206	-0.0232*	-0.0263*	-0.0114
	(0.0124)	(0.0110)	(0.0198)	(0.0114)	(0.0107)	(0.0199)
Customer Tenure (years)				-0.00282***	-0.00286***	-0.00289***
				(0.000335)	(0.000343)	(0.000362)
Size of Wallet ($1,000)				-0.000132**	-0.000125**	-0.000121*
				(0.0000450)	(0.0000437)	(0.0000475)
Share of Wallet (%)				-0.000802***	-0.000790***	-0.000814***
				(0.0000734)	(0.0000745)	(0.0000815)
FICO Score				0.000265***	0.000269***	0.000267***
				(0.0000427)	(0.0000458)	(0.0000480)
Constant	0.221***	0.218***	0.180*	0.0670	0.0773	0.0164
	(0.0530)	(0.0472)	(0.0852)	(0.0559)	(0.0460)	(0.0881)
Observations	40810	38093	32846	40307	37624	32476
R^2	.	.	0.0004	0.0168	0.0200	0.0220
IVs	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)	Rep-Call Type Avg. Sat. (before)	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)	Rep-Call Type Avg. Sat. (before)
First stage partial R^2	0.0206	0.0263	0.0062	0.0204	0.0261	0.0061

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions; and 4) we use the linear probability model here for its flexibility to control a rich set of fixed effects.

Table 5c Testing the Exogeneity of IVs in the Attrition Equation

	Dependent Variable: Attrition	
	(1)	(2)
Satisfaction	-0.0220	-0.0263
	(0.0214)	(0.0168)
Rep Avg. Sat (after)	-0.00166	
	(0.0124)	
Rep Avg. Sat (before)		0.00153
		(0.0114)
Customer Tenure (years)	-0.00287***	-0.00287***
	(0.000350)	(0.000350)
Size of Wallet ($1,000)	-0.000125**	-0.000125**
	(0.0000446)	(0.0000445)
Share of Wallet (%)	-0.000804***	-0.000802***
	(0.0000774)	(0.0000755)
FICO Score	0.000272***	0.000274***
	(0.0000430)	(0.0000477)
Constant	0.0657	0.0693
	(0.0561)	(0.0458)
Observations	36395	36395
R^2	0.0223	0.0207
IVs	Rep. Avg. Sat. (before)	Rep. Avg. Sat. (after)

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions; and 4) we use the linear probability model here for its flexibility to control a rich set of fixed effects.

3.5. The Assignment of Service Reps and the IV Approach

We now explain in more detail the assignment of reps to calls and the extent to which the assignment is independent of the calling customers. For most reps, each of them belongs to a functional group that specializes in handling calls concerning certain card products. Within some of these major groups, a small set of reps are designated to help "high-value" customers. Besides the major groups of reps that focus specific card products, there are several small groups of reps that focus on handling calls concerning some special issues (e.g. fraud) or with special language preferences. The (work) profiles of the reps clarify their qualifications for handling calls regarding certain card products, customers and special service issues.

The automated call routing process that assigns available reps to calls matches rep profiles with the calling customers' status (regular vs. high-value) and the card products in question. A service call first in its queue gets assigned a rep from the designated group once one of them become available. A rep from another group may be assigned to help a customer if no reps from the designated group become available soon enough. Such less preferred assignments are necessary sometimes because longer time waiting in queues also lowers customer satisfaction. Which particular rep from a designated group (or a non-designated group if necessary) first become available and handles a call in the queue is random to, i.e., independent of, the calling customers (in terms of, e.g., their willingness to recommend or attrition in the future).

In our analysis above, we included the fixed effects of card product/call type and the metrics that determine customer value to control for the (non-random) assignment of reps by card products and customer value. In the following, we provide empirical evidence for the assignments of reps being independent of the calling customers, conditioning on the fixed effects of card products and customer value.

First, Table 6a shows the frequencies of the assignment of reps to calls by the profiles of the assigned reps and the card products in question. We are able to identify the rep profile for only a subset of the reps (928 out of a total of 3675 reps) in our sample due to data limitations.[10] The tabulation shows that, with limited exceptions, the reps are assigned to answer calls concerning the card products covered by their functional groups. For example, reps of the "charge and lending" group handle mainly calls regarding charge cards or lending cards (i.e. credit cards). Within some functional groups, some reps are designated to help high-value customers. For example, reps of 'Charge Lending HVCM tier 2' are designated to help high-value customers of charge or lending cards.

Table 6a The Frequencies of Assignment of Reps to Customers by Rep Profiles and Card Products

	Card Products									
Rep profile	ChgLen card 1	ChgLen card 2	ChgLen card 3	ChgLen card 4	ChgLen card 5	Cobrand card 1	Cobrand card 2	Premium card	Other cards	Total
Charge Lending tier 1	618	1827	290	1634	961	100	56	6	697	6189
Charge Lending HVCM tier 2	52	125	21	540	127	18	5	35	71	994
Cobrand tier 1	1	3	0	3	2	637	336	0	157	1139
Cobrand HVCM tier 2	0	1	0	0	0	356	281	0	173	811
Premium tier 3	1	5	0	142	10	5	2	936	24	1125
ISU tier 1	26	60	22	52	23	104	43	0	46	376
ISU tier 3	6	22	1	15	3	38	10	20	18	133
Bilingual tier 3	17	57	14	50	54	209	16	7	34	458
Other profiles	33	145	14	229	64	80	127	154	116	962
Total	754	2245	362	2665	1244	1547	876	1158	1336	12187

Note: 'ChgLen' is short for 'Charge or Lending'; 'HVCM' is short for 'High-Value Customers'.

Table 6b reports the fixed-effect regressions of the skill (measured by our proxies) of the assigned rep on the calling customer's characteristics, controlling for the fixed effects of

[10] The rep identification variable that uniquely identify each rep in our data is missing for many reps in the source data of rep profiles (which is from the same company but maintained by a team different from the one that provided the data that we use in our (main) analysis).

'card product' by 'call type'. The estimates in columns one and four show that the key variables capturing the value of a customer to the firm are positively correlated with the skill (Rep Avg. Sat (before)) of the assigned rep. Nonetheless, the very small R^2 of 0.003 and 0.0028 suggest that the assignments are almost always determined by reps' random availability. Estimates in column four also shows that, besides the customer-value related measures, rep assignments do *not* depend on any customer demographic variables, suggesting no targeted rep assignments beyond those based on customer value. The regressions of the alternative proxies for rep skill confirm the same qualitative findings.

Table 6b The Assignment of Customer Service Reps

	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)
Size of Wallet (\$1,000)	0.000177***	0.000154**	0.000244***	0.000164**	0.000145**	0.000222***
	(0.0000532)	(0.0000480)	(0.0000615)	(0.0000529)	(0.0000484)	(0.0000605)
Share of Wallet (%)	0.000133**	0.000149*	0.0000765	0.000116*	0.000121	0.0000921
	(0.0000479)	(0.0000582)	(0.000115)	(0.0000542)	(0.0000685)	(0.000118)
FICO Score	0.000115***	0.000134***	0.000176**	0.000119***	0.000122***	0.000120
	(0.0000231)	(0.0000266)	(0.0000609)	(0.0000269)	(0.0000325)	(0.0000703)
Customer Tenure (years)				0.0000473	0.0000497	-0.000194
				(0.000185)	(0.000211)	(0.000451)
Female				-0.00299	-0.000195	-0.00704
				(0.00307)	(0.00463)	(0.00837)
Age				0.0000282	-0.00000751	0.000405
				(0.000121)	(0.000151)	(0.000300)
Constant	4.209***	4.139***	4.173***	4.208***	4.151***	4.202***
	(0.0182)	(0.0205)	(0.0449)	(0.0196)	(0.0237)	(0.0515)
Observations	40307	37624	32476	35409	33052	28577
R^2	0.0030	0.0020	0.0008	0.0027	0.0016	0.0006

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of 'card product' by 'call type' are included in all regressions.

Table 6c reports the same regressions as in Table 6b using the subsample for which the rep profile information is available. The estimates show similar correlations between measures related to customer value and the skill level of the assigned rep. To show the random assignments of reps with the same profile, we control for the fixed effects of 'card product' by 'call type' by 'rep profile' in the same regressions as in Table 6c. Table 6d shows that the correlations between customer-value related measures and rep skill proxies become insignificant, both statistically and in magnitude, once we further condition on the profile of the reps. In addition, the R^2 drops to close to zero for all regressions. These

Table 6c The Assignment of Customer Service Reps, the Subsample with Rep Profile

	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)
Size of Wallet ($1,000)	0.0000694	0.0000290	0.000130*	0.0000611	0.0000278	0.000127*
	(0.0000436)	(0.0000540)	(0.0000569)	(0.0000418)	(0.0000569)	(0.0000549)
Share of Wallet (%)	0.000142	0.000161	-0.00000473	0.0000992	0.000176	-0.0000674
	(0.0000728)	(0.0000915)	(0.000203)	(0.0000801)	(0.000107)	(0.000206)
FICO Score	0.000122***	0.000188***	0.000208	0.000151***	0.000193***	0.000239
	(0.0000339)	(0.0000469)	(0.000109)	(0.0000403)	(0.0000526)	(0.000126)
Customer Tenure (years)				-0.000174	-0.000400	0.0000964
				(0.000257)	(0.000411)	(0.000718)
Female				-0.00941	-0.00112	-0.00957
				(0.00585)	(0.00890)	(0.0141)
Age				0.0000439	0.000369	-0.0000824
				(0.000178)	(0.000299)	(0.000526)
Constant	4.224***	4.109***	4.181***	4.209***	4.090***	4.172***
	(0.0245)	(0.0352)	(0.0788)	(0.0288)	(0.0404)	(0.0940)
Observations	11595	11809	9600	10057	10249	8336
R^2	0.0019	0.0015	0.0005	0.0019	0.0014	0.0002

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of 'card product' by 'call type' are included in all regressions.

Table 6d The Assignment of Customer Service Reps, the Subsample with Rep Profile

	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)
Size of Wallet ($1,000)	0.0000147	-0.0000138	0.0000513	0.0000133	-0.00001000	0.0000443
	(0.0000234)	(0.0000162)	(0.0000438)	(0.0000231)	(0.0000155)	(0.0000396)
Share of Wallet (%)	-0.0000612	-0.0000832	-0.000284	-0.0000716	-0.0000490	-0.000395
	(0.0000601)	(0.0000654)	(0.000229)	(0.0000675)	(0.0000700)	(0.000230)
FICO Score	0.0000249	0.0000251	0.000113	0.0000268	-0.00000545	0.000144
	(0.0000276)	(0.0000366)	(0.000113)	(0.0000329)	(0.0000428)	(0.000130)
Customer Tenure (years)				-0.000146	-0.000537	0.000473
				(0.000185)	(0.000275)	(0.000765)
Female				-0.00545	-0.000879	-0.00399
				(0.00443)	(0.00573)	(0.0134)
Age				0.0000138	0.000375	-0.000149
				(0.000122)	(0.000200)	(0.000506)
Constant	4.311***	4.247***	4.271***	4.316***	4.257***	4.264***
	(0.0205)	(0.0271)	(0.0812)	(0.0247)	(0.0312)	(0.0976)
Observations	11595	11809	9600	10057	10249	8336
R^2	-0.0001	-0.0001	0.0001	-0.0001	0.0003	0.0000

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of 'card product' by 'call type' by 'rep profile' are included in all regressions.

results suggest that the correlations we observe in Tables 6b and 6c between customer-value measures and rep skill proxies are only due to the relatively higher skill levels of reps designated to help high-value customers.

One might be concerned that the coefficients of customer-value related measures become insignificant in Table 6d only because there is less variation available for identification after controlling for the additional fixed effects. To address the concern, we report in Table

Table 6e The Assignment of Customer Service Reps, the Subsample with Rep Profile

	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)	Rep Avg. Sat (before)	Rep Avg. Sat (after)	Rep-Call Type Avg. Sat (before)
Size of Wallet ($1,000)	0.0000434	-0.0000122	0.000104*	0.0000344	-0.0000189	0.000121*
	(0.0000423)	(0.0000618)	(0.0000504)	(0.0000417)	(0.0000671)	(0.0000493)
Share of Wallet (%)	0.000126	0.000198	-0.0000922	0.0000733	0.000187	-0.0000394
	(0.0000862)	(0.000112)	(0.000226)	(0.0000994)	(0.000133)	(0.000226)
FICO Score	0.000102*	0.000179**	0.0000793	0.000127**	0.000178**	0.0000698
	(0.0000396)	(0.0000560)	(0.000130)	(0.0000479)	(0.0000576)	(0.000152)
Customer Tenure (years)				-0.000174	-0.000397	-0.000553
				(0.000281)	(0.000444)	(0.000770)
Female				-0.0123	-0.000674	-0.00197
				(0.00703)	(0.0102)	(0.0146)
Age				0.0000147	0.000298	0.000278
				(0.000215)	(0.000379)	(0.000576)
Constant	4.241***	4.115***	4.284***	4.233***	4.106***	4.285***
	(0.0291)	(0.0424)	(0.0931)	(0.0339)	(0.0464)	(0.114)
Observations	11595	11809	9600	10057	10249	8336
R^2	0.0010	0.0015	0.0000	0.0011	0.0012	-0.0002

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of 'card product' by 'call type' by '*Pseudo* rep profile' are included in all regressions, where the 'Pseudo rep profile' is generated through a random permutation of the profiles at the rep level.

6e the fixed-effect regression results that controls for the fixed effects of 'card product' by 'call type' by '*pseudo* rep profile', where the 'pseudo rep profile' is generated by a random permutation of the rep profile at the rep level. The estimates in Table 6e show that the correlations between customer-value related measures and rep skill proxies remain significant if we control for the 'pseudo rep profile' as opposed to the actual rep profile. In addition, the R^2 is also similar to the corresponding ones reported in Table 6c. These additional results together with the ones we presented above suggest that the assignment of reps is indeed random once we further condition on rep profile.

Our discussion above shows that the skill level of the assigned rep is a legitimate IV for customer satisfaction once we include the necessary fixed effects and control variables. Assuming the availability of the relevant control variables for rep assignment is typically not restrictive for firms or researchers using firms' internal data.

3.6. Robustness of the Main Findings

The IV approach introduced in the paper provides a method to obtain a consistent estimate of the causal impact of service satisfaction specific to individual firms (and service activities within firms) by using routine customer survey data and internal data available within firms. The qualitative finding that the effect of satisfaction has been underestimated is likely not limited to our specific data and company, because the attenuation bias caused by the measurement error in satisfaction has not been formally addressed in the past. To obtain the quantitative estimates of satisfaction's causal impact for individual firms, our IV approach can be applied to the data from their own customer satisfaction programs.

A potential issue with our IV estimates of the casual impact of customer satisfaction is that they are based on the sample of customers who called in during the data period *and* responded to the follow-up surveys, instead of all those who called in our data period. The survey response rate is typically about 5%, and thus response bias could be a potential issue. Addressing the response bias will not be a problem for firms, as the selection effect can be controlled for by jointly estimating the RTF (attrition) equation and the binary-choice model for survey responses (c.f. section 24.5 in Greene 2008). Unfortunately, the credit-card issuer that we worked with did not provide us the data on customers who called but did not respond to surveys. We therefore use an indirect approach based on survey response timing to test whether the response bias have a significant effect on our estimates.

The company sends the customer satisfaction survey to every customer who received service, the day after the service encounter, and allows up to two weeks to receive a response. We are able to identify the survey response timing for around half the observations, but not for the rest, in our data, due to how the data were shared with us. Table A2 in the appendix shows that the subsample with information on survey response timing is similar to the full sample in all the summary statistics. Table 7a reports the means of Satisfaction, RTF, Attrition, Age and Customer Tenure by the response timing (i.e., the number of days since the service encounter until the response to the survey). The frequency by the day of response shows that although some customers respond to the survey quickly, most customers respond to the survey only after some delay.

More importantly, the patterns in the Table 7a suggest that the selection effect is limited to the responses within five days after the service encounter. Those who respond within five days show significantly lower average satisfaction level and RTF, but there is little variation in the average satisfaction and RTF across days for those who responded after five days. The attrition rate for those who respond on the second day is one percentage point higher than that for those who respond on most other days; customers who respond earlier are somewhat older than those who respond later; and the tenures of those who respond within three days are somewhat shorter than those who respond later.

Table 7a Descriptive Statistics by Time To Respond to Survey

Days Until Response	Freq.	Satisfaction	RTF	Attrition	Age	Customer Tenure
2	678	3.57	7.42	0.11	57.21	10.93
3	699	3.75	7.81	0.09	56.63	10.52
4	1,153	3.87	7.88	0.10	56.74	11.74
5	1,430	3.93	7.91	0.10	56.25	11.39
6	3,856	4.24	8.43	0.10	56.98	11.35
7	5,755	4.26	8.43	0.10	56.76	11.24
8	3,010	4.24	8.40	0.10	55.77	11.29
9	1,313	4.24	8.33	0.12	55.45	11.72
10	1,344	4.18	8.28	0.10	55.57	10.77
11	315	4.24	8.57	0.08	54.07	11.72
12	200	4.14	8.20	0.10	53.53	9.74
13	172	4.25	8.59	0.08	54.36	9.77
14	51	4.04	8.73	0.08	55.62	11.28

Note: This table is based on the subsample for which the survey response timing is available.

To assess the impact of the sample selection on our results, we compare the estimates of the RTF and attrition equations based on the subsample of customers who respond within five days with the corresponding estimates based on the entire sample (see Tables 7b and 7c). The first two columns in Table 7b report the OLS and IV estimates, respectively, of the RTF equation using the subsample of customers who respond within five days, and the last two columns report the corresponding estimates using the entire sample. Even though the subsample of those who responded within five days is most affected by the selective effect, the OLS and IV estimates based on the subsample are not significantly different from those based on the entire sample. Table 7c shows similar findings regarding the estimates of the attrition equation.[11]

If we view customers who did not respond to the survey as those who delayed their responses to more than 14 days later, the summary statistics in Table 7a suggest that they are likely similar to those who completed the survey in the second week after the service encounter. The estimates in Tables 7b and 7c show that the impact of the sample selection issue on the estimates seems very limited. Taken together, the evidence presented above suggests that our main findings in the previous sections should not be significantly affected by the sample selection problem.

4. The Differential Impact of Satisfaction Across Call Types

In this section, we apply our IV strategy to assess the heterogeneity in the causal impact of service satisfaction on customer loyalty across call types. We first estimate how the

[11] The lack of significance of the coefficient of 'Satisfaction' in column two of Table 7c is likely due to the much smaller size of the subsample.

Table 7b Customer Satisfaction and RTF: Robustness to Selection in Response to Survey

	Dependent Variable: RTF			
	OLS	IV	OLS	IV
Satisfaction	1.223***	1.724***	1.108***	1.941***
	(0.0468)	(0.214)	(0.0333)	(0.186)
Customer Tenure (years)	0.00410	0.00713	0.00420**	0.00474**
	(0.00379)	(0.00415)	(0.00150)	(0.00158)
Size of Wallet ($1,000)	-0.00107*	-0.00102	-0.0000711	-0.000230
	(0.000513)	(0.000655)	(0.000213)	(0.000212)
Share of Wallet (%)	0.00272*	0.00232	0.00300***	0.00282***
	(0.00115)	(0.00134)	(0.000536)	(0.000617)
FICO Score	-0.000514	-0.000687	-0.000163	-0.000509
	(0.000719)	(0.000745)	(0.000302)	(0.000338)
Constant	3.371***	1.569	3.598***	0.404
	(0.541)	(0.801)	(0.279)	(0.760)
Observations	3910	3733	19715	18953
R^2	0.418	0.423	0.275	0.277
# of days until response	≤ 5	≤ 5	≤ 14	≤ 14

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table 7c Customer Satisfaction and Attrition: Robustness to Selection in Response to Survey Selection

	Dependent Variable: Attrition			
	OLS	IV	OLS	IV
Satisfaction	-0.0126**	-0.0323	-0.0111***	-0.0454*
	(0.00383)	(0.0230)	(0.00187)	(0.0181)
Customer Tenure (years)	-0.00321***	-0.00339***	-0.00333***	-0.00342***
	(0.000646)	(0.000692)	(0.000392)	(0.000398)
Size of Wallet ($1,000)	0.0000183	0.0000267	-0.000102	-0.0000960
	(0.000150)	(0.000152)	(0.0000546)	(0.0000538)
Share of Wallet (%)	-0.00118***	-0.00120***	-0.000983***	-0.000978***
	(0.000194)	(0.000200)	(0.000102)	(0.000107)
FICO Score	0.000266**	0.000304***	0.000327***	0.000357***
	(0.0000842)	(0.0000923)	(0.0000577)	(0.0000567)
Constant	0.0447	0.0942	-0.00994	0.112
	(0.0619)	(0.0979)	(0.0380)	(0.0841)
Observations	3910	3733	19715	18953
R^2	0.0276	0.0292	0.0190	0.0178
# of days until response	≤ 5	≤ 5	≤ 14	≤ 14

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions; and 4) we use the linear probability model here for its flexibility to control a rich set of fixed effects.

difficulty of the calls moderates the causal impact of service satisfaction. Then, we separately estimate the impact of service satisfaction for four specific types of calls. The latter exercise allows us to also assess the differential extent to which service satisfaction affects customer loyalty across the four types of calls.

4.1. The Heterogeneity in Satisfaction's Causal Impact across Call Types

The effect of service satisfaction on RTF and loyalty can vary across different types of calls, due to, for example, the differences in the importance of the requests to customers and the extent to which reps are able to satisfy the requests. For the analysis in this subsection, we classify calls into three categories—"hard," "average" and "easy"—based on the average satisfaction ratings for the calls. We also define a fourth category of calls, "cancel card," as these customers are likely the most dissatisfied at the time of call and probably the most difficult to retain. The hard calls include calls to check or request a change in APR and to request an increase in the credit line; and the easy calls include the ones to make a payment, inquire about balance, clarify a unrecognizable charge, check recent charges, replace a lost, stolen card, or inquire about user ID or password. The rest are the average calls, which include, for example, those for questioning a fee and for disputing an inappropriate charge. The hard, easy and "cancel card" calls account for 6%, 36% and 3%, respectively, of all calls in our entire survey sample.

The four categories of calls require different levels of effort by the reps. The easy calls requires the least amount of effort to satisfy the customers. The average calls require some effort by the reps to, for example, follow the appropriate procedures. The hard calls require the most effort by the reps to satisfy the customers. The reps may need to examine a customer's status and make some potentially discretionary decisions to either explain why the customer's requests cannot be granted, or satisfy certain requests within their authority, or even escalate to get his/her managers involved.

We empirically analyze how the impact of customer satisfaction on RTF and attrition varies across the four categories of calls. The results of the RTF regressions, augmented with the interactions of Satisfaction and call category dummies, are reported in Table 8a. The OLS regressions in columns one and four show that, relative to average calls, customer satisfaction of hard (easy) calls has a significantly larger (smaller) positive impact on RTF. In line with the OLS findings, the reduced-form regressions in columns two and five show that, relative to average calls, the impact of rep skill is significantly larger (smaller) for hard (easy) calls. The IV estimates in columns three and six show that customer satisfaction of hard calls indeed has a significantly larger causal impact on RTF, and the impact of customer satisfaction with easy calls is smaller relative to average calls, although the difference for the latter is statistically insignificant. The insignificance of easy calls'

Table 8a Customer Satisfaction and RTF: the Heterogeneity in the Causal Effect across Call Types

	Dependent Variable: RTF					
	OLS	OLS	IV	OLS	OLS	IV
Satisfaction	1.231***		1.793***	1.226***		1.776***
	(0.0307)		(0.0870)	(0.0308)		(0.0881)
Satisfaction×*HardCalls*	0.372***		0.530**	0.379***		0.564***
	(0.0555)		(0.164)	(0.0562)		(0.167)
Satisfaction×*EasyCalls*	-0.248***		-0.0997	-0.249***		-0.105
	(0.0389)		(0.248)	(0.0395)		(0.247)
Satisfaction×*CancelCardCalls*	0.112		0.695	0.129		0.831
	(0.0945)		(0.948)	(0.0927)		(0.958)
Rep Avg. Sat (before)		1.050***			1.027***	
		(0.0923)			(0.0928)	
Rep. Avg. Sat×*HardCalls*		1.031***			1.095***	
		(0.250)			(0.253)	
Rep. Avg. Sat×*EasyCalls*		-0.391**			-0.377**	
		(0.135)			(0.137)	
Rep. Avg. Sat×*CancelCardCalls*		-0.245			-0.223	
		(0.610)			(0.641)	
Customer Tenure (years)				0.00442***	0.00428***	0.00479***
				(0.000950)	(0.00125)	(0.000943)
Size of Wallet ($1,000)				-0.0000826	-0.0000412	-0.000226
				(0.000149)	(0.000232)	(0.000160)
Share of Wallet (%)				0.00258***	0.00319***	0.00217***
				(0.000353)	(0.000403)	(0.000378)
FICO Score				0.0000474	0.000553*	-0.000201
				(0.000200)	(0.000250)	(0.000211)
Constant	3.575***	4.387***	0.821	3.379***	3.814***	0.876*
	(0.0828)	(0.284)	(0.441)	(0.183)	(0.336)	(0.428)
Observations	42337	40810	40810	41814	40307	40307
R^2	0.2881	0.0171	0.1910	0.2895	0.0205	0.1931

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

moderating effect here is not very surprising given that the easy calls' moderating effect is also statistically less significant in the reduced-form regressions in columns two and five.

The results of the augmented attrition regressions are reported in Table 8b. In contrast to the results on RTF, we find that the relative difficulty of the calls do not significantly affect how customer satisfaction impacts attrition. For example, the OLS estimates in columns one and four show there is a relatively smaller impact of customer satisfaction on attrition for easy calls, but the moderating effect is statistically insignificant in both the reduced-form regressions and the IV regressions.

The reduced-form regression in column two of Table 8b shows that rep skill is valuable when dealing with customers calling to cancel a card. The impact of rep skill here is big in its magnitude: a one point increase in Rep Avg. Sat reduces the attrition rate by 9 ppts, statistically significant at the 10% level, more for customers who made cancel-card calls relative to those who made average calls. The corresponding moderating effect is also

Table 8b Customer Satisfaction and Attrition: the Heterogeneity in the Causal Effect across Call Types

	Dependent Variable: Attrition					
	OLS	OLS	IV	OLS	OLS	IV
Satisfaction	-0.0141***		-0.0387*	-0.0130***		-0.0291
	(0.00185)		(0.0169)	(0.00186)		(0.0161)
Satisfaction×$HardCalls$	0.00486		0.0343	0.00310		0.0154
	(0.00557)		(0.0252)	(0.00571)		(0.0237)
Satisfaction×$EasyCalls$	0.00620*		0.0211	0.00562		0.0209
	(0.00300)		(0.0312)	(0.00301)		(0.0301)
Satisfaction×$CancelCardCalls$	-0.0352*		-0.313	-0.0307		-0.178
	(0.0149)		(0.325)	(0.0171)		(0.226)
Rep Avg. Sat (before)		-0.0226*			-0.0168	
		(0.0105)			(0.00996)	
Rep. Avg. Sat×$HardCalls$		0.0188			0.00416	
		(0.0197)			(0.0183)	
Rep. Avg. Sat×$EasyCalls$		0.0158			0.0138	
		(0.0148)			(0.0142)	
Rep. Avg. Sat×$CancelCardCalls$		-0.0910			-0.0479	
		(0.0528)			(0.0461)	
Customer Tenure (years)				-0.00280***	-0.00282***	-0.00281***
				(0.000327)	(0.000337)	(0.000334)
Size of Wallet ($1,000)				-0.000135**	-0.000134**	-0.000129**
				(0.0000450)	(0.0000456)	(0.0000448)
Share of Wallet (%)				-0.000793***	-0.000814***	-0.000791***
				(0.0000704)	(0.0000723)	(0.0000735)
FICO Score				0.000253***	0.000255***	0.000268***
				(0.0000425)	(0.0000430)	(0.0000427)
Constant	0.145***	0.171***	0.255***	0.0245	0.0316	0.0735
	(0.00613)	(0.0307)	(0.0705)	(0.0275)	(0.0438)	(0.0661)
Observations	42337	40810	40810	41814	40307	40307
R^2	0.0028	0.0005	.	0.0232	0.0215	0.0216

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

present in the OLS and IV regressions, though not as statistically significant. The statistical insignificance is possibly a result of the relatively small size of our analysis sample.

The findings above suggest that increasing the satisfaction for the difficult calls, relative for the average calls, is more effective for improving the calling customers' overall RTF; and improving the satisfaction for calls to cancel may be more effective in lowering the calling customers' attrition rate.

4.2. Estimating Satisfaction's Casual Impact by Call Types

The analysis above shows that customer satisfaction with the more difficult calls has larger marginal impact on RTF and customer loyalty. In this subsection we report the analysis by a few representative call types, which shows that customer satisfaction with more difficult calls explains more variation in RTF and attrition, and confirms again that the customer satisfaction is a more influential factor for more difficult calls.

Table 9a shows that rep skills have a larger impact and explain more variation in satisfaction for some types of calls than for other calls. For example, the marginal impact of Rep Avg. Sat (before) is more than two times larger for type 4 calls ("Request a change in your APR") than for type 1 calls ("Inquires about your balance/account/bill"), and it explains eight times more variation in satisfaction for type 4 calls than for type 1 calls. In contrast to Table 9a, Table 9b shows that customer characteristics explain similarly little variation in satisfaction across different call types.

Table 9a Customer Satisfaction and Rep Skill, by Call Types

	Dependent Variable: Satisfaction			
	Type 1 Calls	Type 2 Calls	Type 3 Calls	Type 4 Calls
Rep Avg. Sat (before)	0.434***	0.429***	0.838***	0.933***
	(0.0442)	(0.0750)	(0.109)	(0.0294)
Constant	2.474***	2.601***	0.508	-0.645***
	(0.191)	(0.325)	(0.463)	(0.122)
Observations	4704	5009	2779	1132
R^2	0.0111	0.0135	0.0616	0.0809

Note: 1) Type 1 calls: "Inquire about your balance/account/bill", type 2 calls: "Dispute an inappropriate or incorrect charge", type 3 calls: "Question a fee or charge", type 4 calls: "Request a change in your APR"; 2) Standard errors, clustered at the card product/call type level, in parentheses; 3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 4) the fixed effects of card product/call type are included in all regressions.

Table 9b Customer Satisfaction and Customer Characteristics, by Call Types

	Dependent Variable: Satisfaction			
	Type 1 Calls	Type 2 Calls	Type 3 Calls	Type 4 Calls
Customer Tenure (years)	0.000530	0.00318*	-0.000281	0.00161
	(0.00101)	(0.00118)	(0.00257)	(0.00676)
Size of Wallet ($1,000)	0.000307*	-0.000334	0.000150	-0.000273
	(0.000144)	(0.000348)	(0.0000778)	(0.00224)
Share of Wallet (%)	0.000618*	0.000310	0.00158***	-0.00120
	(0.000263)	(0.000467)	(0.000386)	(0.000948)
FICO Score	0.000745***	0.000603**	0.000770	-0.00125
	(0.000174)	(0.000203)	(0.000768)	(0.00129)
Constant	3.735***	3.946***	3.395***	4.200***
	(0.126)	(0.164)	(0.572)	(1.006)
Observations	4822	5130	2844	1161
R^2	0.0040	0.0046	0.0041	0.0036

Note: 1) Type 1 calls: "Inquire about your balance/account/bill", type 2 calls: "Dispute an inappropriate or incorrect charge", type 3 calls: "Question a fee or charge", type 4 calls: "Request a change in your APR"; 2) Standard errors, clustered at the card product/call type level, in parentheses; 3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table 9c presents the estimates of the RTF equation by call types. The OLS estimates are all smaller than the IV estimates, consistent with our findings using the entire sample.

Satisfaction for type 4 calls has the largest impact on RTF. The finding is consistent with the intuition that the type 4 (about APR) is the most important one of the four. [12] Adding the controls of customer characteristics has little impact on these results (see Table A3 in the Appendix).

Table 9c Customer Satisfaction and RTF, by Call Types

	Dependent Variable: RTF			
	Type 1 Calls	Type 2 Calls	Type 3 Calls	Type 4 Calls
Model 1: OLS				
CM Satisfaction	1.050^{***}	1.152^{***}	1.516^{***}	1.583^{***}
	(0.0504)	(0.0392)	(0.0485)	(0.0398)
Constant	4.184^{***}	3.842^{***}	1.804^{***}	0.682^{***}
	(0.219)	(0.175)	(0.197)	(0.129)
Observations	4868	5147	2861	1163
R^2	0.2174	0.2700	0.4325	0.4386
Model 2: IV Regression				
CM Satisfaction	1.829^{***}	1.400^{***}	1.690^{***}	2.199^{***}
	(0.387)	(0.169)	(0.106)	(0.136)
Constant	0.796	2.736^{***}	1.095^{*}	-1.315^{**}
	(1.681)	(0.753)	(0.429)	(0.439)
Observations	4704	5009	2779	1132
R^2	0.105	0.257	0.427	0.375
Model 3: Reduced form				
Rep Avg. Sat (before)	0.793^{***}	0.601^{***}	1.416^{***}	2.053^{***}
	(0.190)	(0.118)	(0.215)	(0.124)
Constant	5.321^{***}	6.378^{***}	1.954^{*}	-2.733^{***}
	(0.820)	(0.512)	(0.914)	(0.517)
Observations	4704	5009	2779	1132
R^2	0.0073	0.0054	0.0331	0.0683

Note: 1) Type 1 calls: "Inquire about your balance/account/bill", type 2 calls: "Dispute an inappropriate or incorrect charge", type 3 calls: "Question a fee or charge", type 4 calls: "Request a change in your APR"; 2) Standard errors, clustered at the card product/call type level, in parentheses; 3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table 9d show that the IV estimates of the impact of satisfaction on attrition are negative for types 2, 3 and 4 calls, and are statistically significant for types 2 and 4 calls. The R^2 in the regressions for type 4 calls is 0.52% for OLS and 0.22% for the reduced form, both of which are noteworthy for only a single service encounter. In contrast, the impact of satisfaction on attrition for type 1 calls is negative but relatively weaker in both the magnitude and statistical significance according to the OLS estimate, and is insignificant according to the IV estimate. The results are consistent with the calls to "dispute an

[12] A bit surprisingly, the impact of satisfaction on RTF is also quite significant for type 1 calls (balance inquiries) according to the point estimate. This pattern is in contrast to the much smaller effect of rep skill on RTF for type 1 calls than for type 4 calls. The surprisingly large point estimate in the IV regression for type 1 calls is likely a result of sample variance, noting the significantly larger variance of the corresponding estimate.

Table 9d Customer Satisfaction and Attrition in the Following 18 months, by Call Reason

	Dependent Variable: Attrition			
	Type 1 Calls	Type 2 Calls	Type 3 Calls	Type 4 Calls
Model 1: OLS				
CM Satisfaction	-0.0102*	-0.0158***	-0.0154**	-0.0144**
	(0.00367)	(0.00387)	(0.00531)	(0.00389)
Constant	0.121***	0.132***	0.156***	0.139***
	(0.0160)	(0.0173)	(0.0216)	(0.0126)
Observations	4868	5147	2861	1163
R^2	0.0012	0.0032	0.0041	0.0052
Model 2: IV Regression				
CM Satisfaction	0.0267	-0.0570*	-0.0315	-0.0329***
	(0.0467)	(0.0270)	(0.0353)	(0.00551)
Constant	-0.0386	0.316**	0.223	0.200***
	(0.203)	(0.120)	(0.143)	(0.0178)
Observations	4704	5009	2779	1132
Model 3: Reduced form				
Rep Avg. Sat (before)	0.0116	-0.0245*	-0.0264	-0.0307***
	(0.0199)	(0.0111)	(0.0320)	(0.00532)
Constant	0.0274	0.167**	0.207	0.221***
	(0.0859)	(0.0481)	(0.136)	(0.0221)
Observations	4704	5009	2779	1132
R^2	0.0001	0.0006	0.0011	0.0022

Note: 1) Type 1 calls: "Inquire about your balance/account/bill", type 2 calls: "Dispute an inappropriate or incorrect charge", type 3 calls: "Question a fee or charge", type 4 calls: "Request a change in your APR"; 2) Standard errors, clustered at the card product/call type level, in parentheses; 3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 4) the fixed effects of card product/call type are included in all regressions.

inappropriate or incorrect charge" or "request a change in APR" being more important in determining the value of card products to consumers.

5. Managerial Implications

We discuss below, through basic back-of-envelope calculations, the importance of the IV approach for managerial decisions such as investments in customer service and targeted customer service. Our calculations show that biased OLS estimates can lead to a significant underinvestment in customer service.

Let us first examine the implications for calculating the average profit impact of a one-point increase in a customer's satisfaction with her call to customer service. We limit our calculation to the profit impact generated through the casual impact of a customer's satisfaction with her service call experience in Jan 2009 on the probability of the customer canceling her card in the following 18 months. Recall that a one-point increase in satisfaction lowers customer attrition rate in the following 18 months by 1.2 ppts based on the OLS estimates (column 4 Table 5a) and by 2.3 ppts based on the IV estimates (column 6 Table 5a). Thus, we need to calculate the expected profit impact of lowering the probability of a customer canceling her card in the next 18 months by 1.2 ppts, per the OLS estimate, vs. 2.3 ppts, per the IV estimate. We calculate the profit impact per customer as 1.2% (2.3%) times the average Customer Lifetime Value (CLV) (starting at the $19th$ month from Jan 2009).

For the customers in our sample of Jan 2009, the average annual profit is $244; the average annual attrition rate is 7.3%; and the average customer age is 57. Meanwhile, the account tenure in our sample is 4.3, 8.8, 15.7 and 30.6 years at the $50th$, $75th$, $90th$ and $99th$ percentiles, respectively. Given the customer age and account tenures in our sample, we consider three different assumptions on the *upper bound* on the length (in years) of the remaining customer relationship for all customers in the CLV calculations: 5, 10 and 20 years. On top of these uniform upper bounds, we further assume for every customer that the customer tenure at Jan 2009 plus the length of the remaining customer relationship does not exceed 31 years, unless this restriction implies the remaining customer relationship being shorter than one year, in which case we assume it to be just one year. For calculating the CLV, we assume the company's annual discount rate as $\frac{1}{1+0.05}$. To simplify our calculation, we also assume that customer attrition occurs only at the end of each year.

To calculate the average CLV, we first calculate CLV at the individual customer level, using information on the customer-specific annual profit and predicted attrition rate (Fader and Hardie 2010). To provide the estimates under more conservative attrition-rate assumptions, we also calculate CLV, and the corresponding profit impact, assuming each customer's annual attrition rate being 1.25 times the predicted annual attrition rate. We report our results for the various scenarios in Table 10.

The top panel of Table 10 shows results calculated using the predicted annual attrition rate, while the bottom panel shows results calculated using the more conservative annual attrition rate (i.e., 1.25 times the predicted annual attrition rate). Assuming the upper bound on the remaining customer relationship as 10 years, the average CLV is $1624.3; and thus, a one-point increase in the satisfaction with a customer service call increases the company's profit, on average, by $19.2 according to the OLS estimates and $37.6 according to the IV estimates. Under the alternative assumption of the upper bound being 5 or 20 years for the remaining customer relationship, the corresponding profit impact calculated using the IV estimates is also about two times larger than the one calculated using the OLS

Table 10 The Profit Impact of Customer Satisfaction with a Single Service Call ($)

	Upper Bounds on the Remaining Customer Relationship		
	5 years	10 years	20 years
Predicted attrition rate, average annual profit per CM: $243.7			
Life-time value	1021.9	1624.3	2185.6
The impact of a one-point increase in satisfaction (OLS)	12.1	19.2	25.9
The impact of a one-point increase in satisfaction (IV)	23.7	37.6	50.6
The impact of a 0.4-point increase in satisfaction (OLS)	4.8	7.7	10.4
The impact of a 0.4-point increase in satisfaction (IV)	9.5	15.1	20.3
1.25*Predicted attrition rate, average annual profit per CM: $243.7			
Life-time value	1000.4	1563.3	2064.6
The impact of a one-point increase in satisfaction (OLS)	11.8	18.5	24.4
The impact of a one-point increase in satisfaction (IV)	23.2	36.2	47.8
The impact of a 0.4-point increase in satisfaction (OLS)	4.7	7.4	9.8
The impact of a 0.4-point increase in satisfaction (IV)	9.3	14.5	20.3

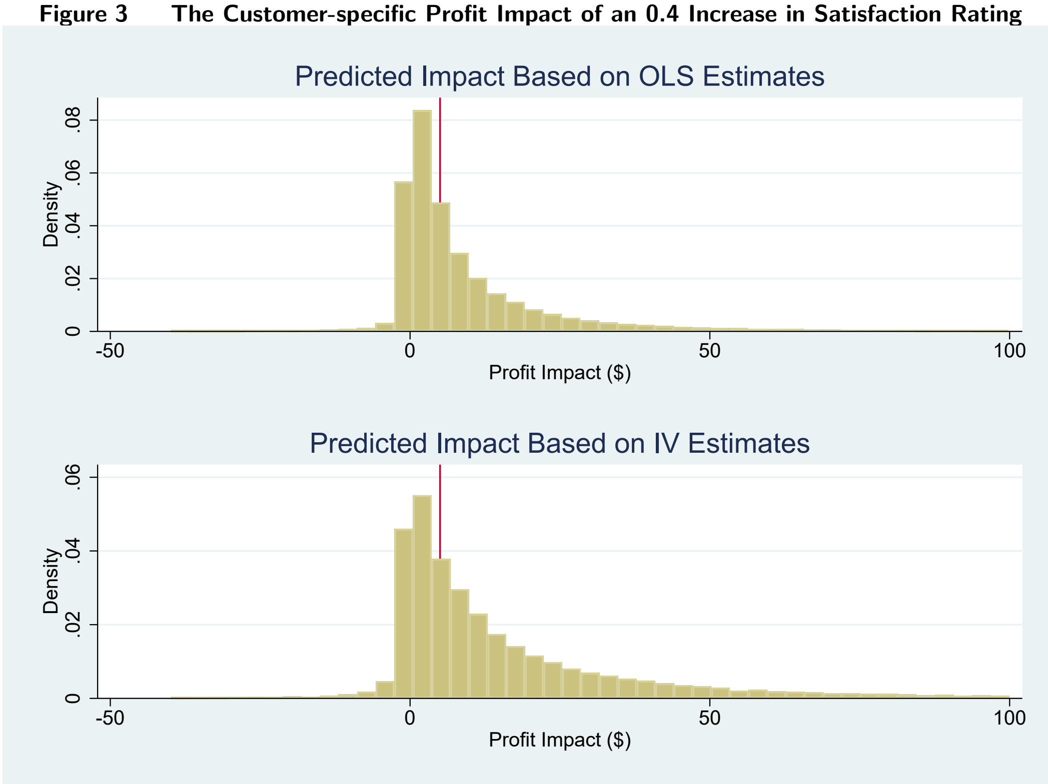
estimates. The bottom panel shows that assuming the more conservative annual attrition rate leads to very similar results.

To provide estimates that are directly relevant to managerial decisions, we also need to know how big an increase in service satisfaction is feasible in practice. The 5th and 95th percentiles of Rep Avg. Sat (before) among all the reps in our 2008-2009 survey sample are respectively 3.9 and 4.6. According to the estimates in column four of Table 3, increasing the call-handling rep's skill level, as measured by our proxy "Rep Avg. Sat (before)", by 0.7 leads to a 0.4-point increase in service satisfaction on average. We report the profit impact of a 0.4-point increase in customer satisfaction in the last two rows of the two panels in Table 10. For example, assuming an upper bound of 10 years of remaining customer relationship with the (1.25 times) predicted annual attrition rate, a 0.4-point in the satisfaction with a customer service call causes an increase in the company's profit by $7.7 ($7.4) and $15.1 ($14.5) on average, according to the OLS and IV estimates respectively.

We make two observations from the above results. First, given the feasibility of an increase of 0.4 in the average customer satisfaction, basing managerial decisions on the OLS estimates can lead to an *underinvestment* in customer service. Second, the company may find it profitable to increase customer satisfaction by improving the skill level of its customer-service workforce. This observation follows given our profit impact estimates above and the fact that the cost of increasing the customer satisfaction by 0.4 point is unlikely to significantly exceed $3 per call.[13] The company may improve the skill level

[13] Suppose the difference in the annual total compensation between a 95th percentile rep and a 5th percentile rep is $36,000 (which is likely an overestimate given the average annual base pay of customer service representatives ranging

Figure 3 The Customer-specific Profit Impact of an 0.4 Increase in Satisfaction Rating



Note: The unit of observation is a customer.

of its customer-service workforce by more effectively recruiting and retaining reps with higher-skill levels. Another possible approach is to provide more training and/or incentives for learning/coaching among peers to improve servicing skills.

The IV and OLS estimates can also lead to very different decisions on the group of customers that the firm may target with premium service. The profit impact of service satisfaction varies across customers due to variation in the customer-level profit and attrition rate. For example, the median and 75th percentile annual profit per customer in our sample are \$142 and \$402, respectively; and the attrition rate for customers calling to request a change in the APR is 9.4%, while that for those asking why a charge was denied is only 4%.

We report in Figure 3 the distribution of the estimated customer-level profit impact assuming the remaining customer relationship to be at most 10 years. Suppose we want to target the higher service quality at customers for whom the profit impact is higher than, for example, \$5 per call (indicated by the vertical line in Figure 3). Then, the OLS-based

from \$23,000 to \$39,000 across companies per glassdoor.com). Each rep in the company answers around 1000 calls per month. Then, the difference in the cost per call between a 5th percentile and a 95th percentile is around \$3.

and IV-based estimates suggest 50% and 63% of the customers in our sample should be targeted with the higher-quality service, respectively. The bias in the OLS estimates would thus lead to significantly fewer customers being targeted with premium customer service.

6. Conclusion

This paper introduces a framework to measure the causal relationship between service satisfaction and loyalty (stated and behavioral) using routine cross-sectional data of surveys conducted by firms after service encounters. The framework addresses multiple sources of bias in measuring the relationship—common methods, measurement error and omitted variables—and allows us to estimate the unbiased magnitude of the relationship. The framework exploits the fact that service employees are typically assigned to customers based on the employees' availability, which is independent of the customers, and thus can be used by firms interested in understanding how investing in service quality and satisfaction provides returns in the form of increased loyalty (RTF and customer retention). The method is robust and can be applied even when the rep assignment depends on observable customer characteristics (e.g., when more important customers are assigned to a division of reps of higher skills), as long as we are able to control for such nonrandom elements in the service assignment process. The estimated causal effect of service satisfaction is also robust to observable customer characteristics (such as customer tenure) being endogenous, as long as the random assignment of service reps is independent of such observable characteristics.[14]

Our results show that the combination of attenuation bias due to "measurement error" in customer satisfaction and (the unsigned nature of) omitted variable bias overwhelms the inflationary bias caused by the common-method problem that is common in research based on survey data. Overall, we show that the literature has thus far underestimated the effect of satisfaction on loyalty intent. When we measure loyalty by a behavioral metric of retention, we find that the link between satisfaction and behavioral loyalty has been underestimated even more relative to the IV estimates. Our results suggest that the ROI on investments in service satisfaction may be grossly underestimated at many firms. Our results also show that the impact of service satisfaction on the stated loyalty is larger for the more difficult/important calls, suggesting the additional value in improving the service quality for such calls.

[14] The IV estimate is not affected when we add additional controls that are independent of the IV.

We hope the current research not only offers a practical approach to evaluate and understand the benefits of investments in customer satisfaction, but helps improve the customer experience at many firms. For academics, the approach potentially provides a way to reconcile the substantial differences in the estimates of the relationship across various studies in past research.

References

Anderson EW, Fornell C, Mazvancheryl SK (2004) Customer satisfaction and shareholder value. *Journal of marketing* 68(4):172–185.

Büschken J, Otter T, Allenby GM (2013) The dimensionality of customer satisfaction survey responses and implications for driver analysis. *Marketing Science* 32(4):533–553.

Cronin JJ, Taylor SA (1992) Measuring service quality: a reexamination and extension. *The journal of marketing* 55–68.

Fader PS, Hardie BG (2010) Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science* 29(1):85–93.

Greene WH (2008) *Econometric analysis* (Pearson).

Heskett JL, Sasser WE (2010) The service profit chain. *Handbook of service science*, 19–29 (Springer).

Kamakura WA (2010) Common methods bias. *Wiley International Encyclopedia of Marketing* .

Kaplan J (2016) The inventor of customer satisfaction surveys is sick of them, too. Bloomberg News, May 5.

Kumar V, Dalla Pozza I, Ganesh J (2013) Revisiting the satisfaction–loyalty relationship: empirical generalizations and directions for future research. *Journal of Retailing* 89(3):246–262.

Menezes AM, Serbin J (1991) *Xerox Corporation: The Customer Satisfaction Program* (Harvard Business School Case Services).

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

Parasuraman A, Zeithaml VA, Berry LL (1985) A conceptual model of service quality and its implications for future research. *the Journal of Marketing* 41–50.

Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP (2003) Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology* 88(5):879.

Reichheld FF (2003) The one number you need to grow. *Harvard business review* 81(12):46–55.

Reichheld FF, Covey SR (2006) *The ultimate question: Driving good profits and true growth* (Harvard Business School Press Boston, MA).

Shankar V, Smith AK, Rangaswamy A (2003) Customer satisfaction and loyalty in online and offline environments. *International journal of research in marketing* 20(2):153–175.

Staiger D, Stock J (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3):557–586.

Wooldridge JM (1995) Score diagnostics for linear models estimated by two stage least squares. *Advances in econometrics and quantitative economics: Essays in honor of Professor CR Rao* 66–87.

Zahorik AJ, Rust RT (1992) Modeling the impact of service quality on profitability: a review. *Advances in services marketing and management* 1(1):247–76.

Zeithaml VA, Berry LL, Parasuraman A (1996) The behavioral consequences of service quality. *the Journal of Marketing* 31–46.

Appendix

Table A1 Satisfaction with Service and Rep Skill

	Dependent Variable: Satisfaction					
	(1)	(2)	(3)	(4)	(5)	(6)
Rep Avg. Sat (before)	0.570***			0.557***		
	(0.0460)			(0.0463)		
Rep Avg. Sat (after)		0.520***			0.525***	
		(0.0311)			(0.0306)	
Rep-Call Type Avg. Sat (before)			0.121***			0.122***
			(0.0146)			(0.0142)
Customer Tenure (years)	-0.000471	-0.000112	-0.000340	-0.000917	-0.000638	-0.000582
	(0.000600)	(0.000603)	(0.000689)	(0.000616)	(0.000641)	(0.000712)
Size of Wallet ($1,000)	0.000101	0.000112	0.0000875	0.0000700	0.0000644	0.0000562
	(0.000102)	(0.000105)	(0.000111)	(0.000105)	(0.000107)	(0.000115)
Share of Wallet (%)	0.000550***	0.000406*	0.000665***	0.000465**	0.000335	0.000512**
	(0.000148)	(0.000158)	(0.000175)	(0.000173)	(0.000184)	(0.000198)
FICO Score	0.000423***	0.000466***	0.000441***	0.000433***	0.000450**	0.000452***
	(0.000105)	(0.000116)	(0.000110)	(0.000126)	(0.000143)	(0.000130)
Female	0.0294*	0.0268*	0.0325*	0.0257*	0.0222	0.0287*
	(0.0118)	(0.0120)	(0.0135)	(0.0120)	(0.0125)	(0.0135)
Age				-0.0000425	-0.000131	-0.000213
				(0.000360)	(0.000364)	(0.000405)
Constant	1.465***	1.687***	3.398***	1.544***	1.717***	3.427***
	(0.193)	(0.144)	(0.104)	(0.178)	(0.133)	(0.111)
N	40307	37624	32476	35409	33052	28577
R^2	0.0222	0.0278	0.00797	0.0214	0.0282	0.00768
(Incremental) F statistics	162.2	303.1	70.2	135.1	127.3	135.6

Note: 1) Standard errors, clustered at the card product/call type level, in parentheses; 2) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.

Table A2 Summary statistics, the Subsample with Survey Response Timing Information

Variable	Mean	Std. Dev.	Min.	Max.	N
CM Satisfaction	4.16	1.06	1	5	19981
Recommend	8.27	2.45	1	10	19981
Customer Tenure (years)	11.25	10.72	0	51	19981
Size of Wallet ($1,000)	35.56	76.62	0	6285.13	19981
Share of Wallet (%)	54.9	34.81	0	100	19888
FICO Score	757.08	61.4	439	997	19778
Customer attrition within 18 months	0.1	0.3	0	1	19981
Rep Avg. Sat (before)	4.31	0.26	2	5	19207
Rep-Call Type Avg. Sat (before)	4.32	0.65	1	5	15636
Rep Avg. Sat (after)	4.25	0.32	1.83	5	19835

Note: 1) The unit of observation is a call with survey result in our sample of Jan 2009; 2) Rep Avg. Sat is the average satisfaction rating of the rep handling a call.

Table A3 Customer Satisfaction and RTF, by Call Types

	Dependent Variable: RTF			
	Type 1 Calls	Type 2 Calls	Type 3 Calls	Type 4 Calls
CM Satisfaction	0.0374	-0.0357	-0.0271	-0.0377***
	(0.0435)	(0.0274)	(0.0322)	(0.00712)
Customer Tenure (years)	-0.00263*	-0.00126	-0.00300*	-0.00248*
	(0.00108)	(0.000926)	(0.00127)	(0.00104)
Size of Wallet ($1,000)	-0.000362***	0.0000212	-0.0000464	-0.000485
	(0.0000869)	(0.0000715)	(0.0000442)	(0.000289)
Share of Wallet (%)	-0.000760***	-0.000811***	-0.000691***	-0.000307*
	(0.000198)	(0.000215)	(0.000160)	(0.000127)
FICO Score	0.000249	0.000155	0.000188	-0.000167
	(0.000139)	(0.000108)	(0.000145)	(0.000205)
Constant	-0.190	0.166	0.124	0.386*
	(0.213)	(0.139)	(0.185)	(0.168)
N	4659	4992	2764	1130
R^2	.	0.0114	0.0165	0.000771

Note: 1) Type 1 calls: "Inquire about your balance/account/bill", type 2 calls: "Dispute an inappropriate or incorrect charge", type 3 calls: "Question a fee or charge", type 4 calls: "Request a change in your APR"; 2) Standard errors, clustered at the card product/call type level, in parentheses; 3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 3) the fixed effects of card product/call type are included in all regressions.