Marketing Science

Rejoinder—Customer Satisfaction-Based Mispricing: Issues and Misconceptions

Robert Jacobson, Natalie Mizik,

Please scroll down for article—it is on subsequent pages

Rejoinder

Customer Satisfaction-Based Mispricing: Issues and Misconceptions

Robert Jacobson

Diogenes Consulting, Seattle, Washington 98109, bob.jacobson@diogenesconsulting.com

Natalie Mizik

Graduate School of Business, Columbia University, New York, New York 10027,
nm2079@columbia.edu

We appreciate the opportunity to respond to the commentaries and additional analyses by Fornell et al. [Fornell, C., S. Mithas, F. V. Morgeson III. 2009a. The economic and statistical significance of stock returns on customer satisfaction. *Marketing Sci.* **28**(5) 820–825] and Ittner et al. [Ittner, C., D. Larcker, D. Taylor. 2009. The stock market's pricing of customer satisfaction. *Marketing Sci.* **25**(5) 826–835]. Both studies have multiple theoretical and econometric limitations that challenge the validity of their arguments and findings (e.g., neither study allows for time-varying risk factor loadings in their assessments of mispricing although the composition of firms in their analyzed portfolios changes over time, Fornell et al. mischaracterize the efficient markets hypothesis, and Ittner et al. do not use standard panel data econometric methods and models). Generalizations about customer satisfaction, like any other construct, should be assessed by appropriate econometric methods and should withstand rigorous scrutiny. We believe an open, frank dialogue can help clear up misconceptions, air central issues, and advance better understanding of methods and analyses for assessing the financial market implications of marketing metrics such as customer satisfaction.

Key words: customer satisfaction; mispricing; value relevance
History: Received: July 17, 2009; accepted: August 5, 2009.

1. Introduction

Before responding to the Fornell et al. (2009a) and Ittner et al. (2009) (FMM2009 and ILT2009, respectively) comments, allow us to first restate what we examine and conclude in Jacobson and Mizik (2009b) (hereafter referred to as JM2009b). JM2009b was motivated by the rudimentary question: What are the conditions that explain when and why customer satisfaction might be mispriced by the financial markets? We could not fully answer this question because we find so little evidence of customer satisfaction-based mispricing. We find that (i) analyses allowing for time-varying risk factor loadings show little evidence to suggest the existence of widespread satisfaction-based mispricing (see Table 3 of JM2009b), and (ii) disaggregate analyses highlight a small number of firms in the computer and Internet sector as a major source of earlier mispricing findings, suggesting that evidence of potential mispricing is limited to this small group of firms.

Unfortunately, FMM2009 obfuscate both considerations; e.g., they provide no substantive discussion of our Table 3 findings and their new empirical analysis neglects both of the above issues. In contrast to FMM2009, we argue the following:

• Researchers need to engage in formal hypothesis testing. Statistical significance (or the lack thereof) of empirical findings is central to marketing science. Claims should not be based on statistically insignificant coefficients and more-likely-than-not criteria. It is tests with correct type I errors that distinguish marketing science-based generalizations from both unsubstantiated assertions and from false claims;

• Although FMM2009 are confused about power considerations (e.g., they misinterpret Kothari and Warner 2001 by mistaking "rejection frequencies" as "failure to reject frequencies"), we agree that statistical power is an important consideration and discuss a number of steps that can enhance power in mispricing studies;

• Particularly when portfolios are rebalanced over time or the period of study involves a lengthy timeframe, time-varying risk factor loadings should be an element of the expected return model. If the firms included in a portfolio change over time, then the risk characteristics of the portfolio can be expected

to change over time as well, thus rendering time-invariant risk model tests of mispricing invalid (Ang and Kristensen 2009); and

• Disaggregate analyses are important and can help identify the sources of the aggregate-level effects.

ILT2009 seek to present a broader analysis of the financial market pricing (as opposed to just mispricing) of customer satisfaction. The entire value implications of customer satisfaction involve both the value relevance (i.e., contemporaneous pricing) and the mispricing (i.e., future-term systematic stock price adjustment) of customer satisfaction. Our differences with ILT2009 relate primarily to how these types of assessments should be carried out. ILT2009 overlook some limitations in their modeling approaches. Furthermore, some of their claims concerning what methods are typical in the accounting and finance literature are at odds with what is actually done in this literature. In contrast to ILT2009, we argue the following:

• Response modeling (Jacobson and Mizik 2009a) is an advantageous means to directly assess the value relevance of a metric. ILT2009 analyses of lagged American Customer Satisfaction Index (ACSI) effects on accounting performance measures are based on models with questionable properties and their event study analyses have limitations (e.g., downward bias associated with measurement error in the unanticipated ACSI measure they use) related to the fact that market participants will update their expectation of customer satisfaction for a firm throughout the year and not just at the ACSI announcement date (Kothari 2001);

• Mispricing studies are orthogonal to value-relevance studies. The most direct tests of mispricing are martingale assessments that investigate whether information available to investors is predictive of future-term abnormal returns (LeRoy 1989). Value relevance is neither a sufficient nor necessary condition for mispricing;

• Mispricing can be based on changes, levels, a combination of changes and levels, or any set of transformations of the available information set. Counter to ILT2009's assertions, an extensive research stream in accounting and finance assessing potential mispricing (e.g., Chan et al. 2001, Edmans 2009) has used portfolio formation criteria based on autocorrelated-level variables, not just unanticipated changes of the metric; and

• ILT2009 tests of mispricing involve a lengthy time period and portfolios whose composition varies over time. As such, their analyses need to allow risk factor loadings to vary intertemporally (Fama 1998).

2. Mispricing vs. Value Relevance

It is useful to distinguish assessments of mispricing from assessments of a metric's value relevance.

Analysis of value relevance assesses the extent to which a metric contains information that has profit implications the financial markets recognize. At their most rudimentary level, analyses of value relevance involve an assessment of the correlation between the unanticipated component of a metric X and *contemporaneous* risk-adjusted stock return. Analysis of mispricing assesses the extent to which the financial markets fail to react to information that has long-term profit implications or overreact to information that does not have long-term profit implications. If financial markets exhibit a delayed response to a metric, the metric is mispriced. At their most rudimentary level, tests of mispricing involve an assessment of the correlation of the metric X (or information contained in the series, such as the first difference in X) with *future-period* abnormal stock return. Because risk-adjusted stock return is not autocorrelated, assessments of value relevance are orthogonal to assessments of mispricing.

Thus, as we discuss on page 818 of JM2009b, findings on mispricing have no implications with respect to value relevance. Rather, our analysis in JM2009b is limited to mispricing, where we fail to find statistically significant evidence of widespread satisfaction-based mispricing.

Indeed, we would argue that empirical evidence shows that customer satisfaction and its components can have profit implications and influence financial market outcomes. For example, the event study of Nayyar (1995) reports that the financial markets value the announcements of customer service increases positively and the decreases negatively. It is reasonable to expect a contemporaneous financial market response to changes in customer satisfaction that have profit implications. In Jacobson and Mizik (2009a), we undertake and report analyses of value relevance. We find that the ACSI metric has value relevance (i.e., changes in the metric exhibit a contemporaneous bivariate association with risk-adjusted return) but not *incremental* value relevance (i.e., changes in the metric are not significantly related to contemporaneous risk-adjusted stock return when changes in accounting variables—in particular, changes in return on assets (ROA)—are also included in the model).

3. FMM2009: Should Generalizations About Customer Satisfaction Be Subject to Rigorous Analysis and Scrutiny?

FMM2009 object to findings that challenge or raise questions about their previous work (see also Fornell et al. 2009b). FMM2009 are averse to analyses assessing the sensitivity of customer satisfaction-based mispricing to alternative models of expected return (e.g.,

allowing for time-varying risk factor loadings), analyses seeking to examine potential differential effects across contexts (e.g., engaging in disaggregate analyses), and even analyses making use of standard financial market methods for assessing mispricing (e.g., calendar-time portfolio analysis). Nonetheless, this is what is required for establishing the validity of findings of mispricing. The lack of statistically significant abnormal returns in, for example, analyses where the expected return model allows for time-varying risk factor loadings calls into question the legitimacy of conclusions and assertions of widespread financial market mispricing of satisfaction. At the very least, it suggests the need for further more detailed and rigorous analyses. The development of marketing science requires that claims about customer satisfaction not be given a free pass, but rather, like every other construct, withstand scrutiny.

Rather than addressing the issues we raise, FMM2009 ignore our key findings (i.e., Table 3 of JM2009b). Furthermore, they make erroneous assertions that misrepresent and misinterpret not only aspects of our analyses but other studies and financial market theory as well.

3.1. Lack of Evidence of Widespread Satisfaction-Based Mispricing

Our analyses that allow for the expected return model to have time-varying risk factor loadings (Table 3 of JM2009b) provide a challenge to claims that widespread satisfaction-based mispricing exists. In the aggregate analysis, none of the portfolios exhibits statistically significant abnormal returns. In the non-public utility/noncomputer and non-Internet sample, portfolio 2 produces higher abnormal returns than the top satisfaction portfolio. Even for computer and Internet firms where the top portfolio generates abnormal returns significant at the 10% level, the returns are not statistically different from portfolio groupings of lower satisfaction firms.

FMM2009 have little to say about our Table 3 results. Instead, they tangentially note (p. 822) that Lewellen and Nagel (2006) report "that betas do not vary enough to explain asset-pricing anomalies." Their statement mischaracterizes the Lewellen and Nagel findings. What Lewellen and Nagel (2006) actually report is that the conditionally estimated market model cannot dissipate the role of other standard risk factors such as momentum and the value premium (i.e., the other risk covariates included in our models). As such, this FMM2009 comment has no relevance for our study and the results we report in Table 3. We estimate time-varying multifactor risk models; that is, as we show in Equation (4), our risk model is not limited to the market model but encompasses the other standard risk factors as well. We reference Lewellen

and Nagel (2006) merely to highlight their use of high-frequency data in order to estimate short-window regressions (as opposed to, for example, instrumenting factor loadings with macroeconomic and firm-specific variables) to allow for time-varying risk factor loadings. A host of other studies (e.g., Fama and French 1997) similarly advocate incorporating time-varying risk factor loadings. Ang and Kristensen (2009) note that

> there is overwhelming evidence that factor loadings, especially for the standard CAPM and Fama and French (1993) models, vary substantially over time even at the portfolio level. ... The time variation in factor loadings distorts the standard factor model tests, which assume constant betas, for whether the alphas are equal to zero and, thus, renders traditional statistical inference for the validity of a factor model to be possibly misleading in the presence of time-varying factor loadings.

As reported in Table 3, panel A, we fail to find widespread evidence of satisfaction-based mispricing in analyses based on calendar-time portfolios with time-varying risk factor loadings. The same is true for analyses based on the use of a multifactor "rolling-window" model as a means of allowing for time-varying risk factor loadings (Table 3, panel B). FMM2009's decision not to address the findings in our Table 3 is unfortunate. These findings run counter to and even render moot many of the arguments FMM2009 raise about the Table 2 results, which we include primarily to show that we can replicate previous findings—in particular, Aksoy et al. (2008). For example, contrary to the FMM2009 assertion, even when using one-tail tests, we fail to find widespread evidence of mispricing. Only by undertaking disaggregate analysis can we find some potential evidence of mispricing for a small subsample of computer and Internet firms.

3.2. FMM Critiques

Contrary to FMM2009 claims, which they repeat in Fornell et al. (2009b), we do not selectively trim or truncate the data prior to aggregate analysis, nor do we trim or truncate the right-hand side of the distribution at any point of our analyses. Rather, as is commonly done, we estimate an aggregate model on all data and also undertake disaggregate analyses on three all-inclusive subgroupings. The literature, highlighted by Pastor and Veronesi (2007), as to why firms that are part of a technological revolution may have properties that differ from other firms is very well established, and the potential difference between public utilities and other firms is similarly widely acknowledged. A Chow test (Chow 1960) confirms that computer and Internet firms have characteristics different from other firms in the sample

and, as such, warrant disaggregate estimation. The FMM2009 criticism of disaggregate analysis is not only unjustified but is also surprising considering that Fornell et al. (2006) and Anderson, Fornell, and Rust (1997), for example, engage in disaggregate analysis (i.e., separating manufacturing from service firms). FMM2009 articulate neither an econometric nor theoretical rationale for why their use of disaggregate analysis (e.g., based on service versus manufacturing firms) is appropriate but our use of disaggregate analysis (e.g., based on computer and Internet firms versus other firms) is not.

FMM2009 advance studies that support their conclusions and are dismissive of studies challenging their conclusions. For example, FMM2009 advance the study of Aksoy et al. (2008) but dismiss the approach of O'Sullivan et al. (2009). However, both studies make use of the same calendar-time portfolio method. These studies differ not so much in methods but rather on the basis by which portfolios are formed. We would advocate a more consistent approach to critiquing empirical work—namely, one that is independent of the particular outcome reported—as a better means of evaluating previous research.

Furthermore, a number of FMM2009 statements are in error. For example, counter to FMM2009 statements (which are a repetition of what they state in Fornell et al. 2009b) claiming that under the efficient market hypotheses the risk "predictor variables must completely account for the return variance" (Fornell et al. 2009b, p. 162), the efficient market hypothesis does not require that "risk factors explain all variation in returns" (FMM2009, p. 822). The efficient market hypothesis allows for, and indeed requires, investors to respond contemporaneously to factors changing current and expected future cash flows. Unfortunately, these incorrect assertions are not isolated instances of FMM2009 misstatements.

3.3. Enhancing the Power of Mispricing Tests

Another example of FMM2009's erroneous statements involves their misinterpretation of Kothari and Warner's (2001) "rejection frequencies" with "failure to reject frequencies." FMM2009 assert, as they do in Fornell et al. (2009b), that the simulation results of Kothari and Warner (2001) show that abnormal annual stock returns "as high as 3% are not detected almost 70% of the time" (FMM2009, p. 822). As shown in Table II, Panel C of Kothari and Warner (2001, p. 1995), for the 4-factor Carhart model in their simulation involving 36 monthly returns with a 3% annual abnormal performance, the rejection frequency using two-tailed tests at the 5% significance level is 69%. That is, the simulation results show that the percentage of time an abnormal return *is* detected in this context is almost 70%—*the exact opposite* of what

FMM2009 assert.[1] Furthermore, FMM2009 also fail to note that Kothari and Warner (2001) show how rejection frequencies increase as the time horizon of the study is extended beyond 36 monthly observations (which is the case in all mispricing studies involving ACSI).

Despite this misinterpretation, we, and most others who study potential mispricing, share FMM2009's concerns with type II errors. For example, Kothari and Warner (2007, p. 23) note that mispricing tests can have "low power against economically interesting null hypotheses." Nonetheless, concern over power does not mean conclusions with respect to mispricing should be based on analysis absent tests of statistical significance or that they should be based on a more-likely-than-not threshold. Tests with correct type I errors are central to marketing science and distinguish marketing generalizations from both unsubstantiated assertions and false claims. Kothari and Warner (2007, p. 21) note that absent proper controls (e.g., for expected return), past research "demonstrates how easy it is to conclude there is abnormal performance when none exists."

Furthermore, statistical power is not invariant but depends on the research framework. For example, power depends on the number of time-series observations used in the analysis. This involves not only the number of firms included in a portfolio but also the length of time the return to the portfolio is assessed. The Kothari and Warner (2001) simulation results FMM2009 highlight (albeit erroneously interpret) are based on only 36 monthly observations. Clearly (as shown in Table VI of Kothari and Warner 2001),

---

[1] We wonder why the FMM2009's statements are so much at odds from what Kothari and Warner (2001) actually report. We considered the possibility that perhaps rather than misinterpreting the simulation findings, FMM2009 instead ignored findings reported in Table II, panel C (p. 1995) and made use of the simulation results that give greater weight to small capitalization stocks (Table III, panel B, p. 1999). However, this possibility is highly unlikely because (i) it is clearly inappropriate to compare the ASCI sample of firms (which are generally very large) to a sample with greater weight given to small capitalization stocks, (ii) the simulation results for small capitalization stocks have lower rejection frequencies than what FMM2009 state, and (iii) the only table in Kothari and Warner (2001) reporting a two-tailed test "(that J&M rely on)" (FMM2009, p. 822) is Table II, panel C. In any event, reconciling FMM2009's claim (p. 822) that Kothari and Warner (2001) report "returns even as high as 3% are not detected almost 70% of the time" with the simulation results Kothari and Warner (2001) actually report is difficult.

A key limitation to the applicability of Kothari and Warner (2001) to our analysis is that they do not allow for time-varying risk factor loadings even though in their simulations, they rebalance portfolios 100% every 12 months. Although we can make some speculations, e.g., confidence intervals would be more correctly sized, we do not know the extent to which their findings would be affected by analyses taking into account the changing risk characteristics of their portfolios.

improving power by making use of more observations is advantageous. Because undertaking analyses with more recent data increases sample size, we were glad to see Table 1 of FMM2009, where they engaged in hypothesis testing (which they did not do in the assessment of mispricing in Fornell et al. 2006) that incorporated more recent data through January 2009. Given their focus on statistical power and sample size, however, we are left to wonder why they did not report analyses utilizing the first years of the ACSI survey (Fornell et al. 2006 present results beginning in February 18, 1997, whereas FMM2009 data analysis begins May 1, 2000). They exclude approximately 25% of available observations. It appears from Figures 2 and 3 in Fornell et al. (2006, p. 9) that including the entire data period, as well as allowing for, for example, time-varying risk factor loadings, might impact their findings.

Low power can also be addressed by moving from portfolio-level analysis to firm-level analysis. Although this move would seem an obvious conclusion from panel data econometrics, some researchers still object to undertaking anything but portfolio-level estimation. Nonetheless, the panel data firm-level random effects (i.e., generalized least squares (GLS)) estimator is more efficient than the portfolio estimator (Hsiao 1986), which is simply a panel data "between-period" estimator. A common misperception is that estimation based on portfolios yields more accurate estimates than firm-level analysis because it is based on averages across many stocks. In fact, not only will the variance of the random-effects estimator *always* be smaller than the variance of the between (portfolio-level) estimator asymptotically, but even for relatively small and moderate sample sizes, the feasible GLS estimator will be more efficient than alternative panel data estimators (Taylor 1980). Furthermore, although efficiency comparison among other panel data estimators depends on the particular characteristics of a data set, Biørn (2001) reports other panel data estimators (e.g., the pooled ordinary least squares estimator) are often more efficient than the between-period (i.e., portfolio) estimator. The advantage of the portfolio (i.e., between) estimator is that it circumvents the issue of cross-sectional dependency. However, other approaches (e.g., working with a matched firm or utilizing cluster robust standard errors) can be used to address these concerns without sacrificing/collapsing firm-specific variation into a portfolio.

Kothari and Warner (2007, p. 23) do not advocate the random-effects estimator per se, but do view firm-level analysis and portfolio analysis as viable alternatives and state, "Despite an extensive literature, there is still no clear winner in a horse race." As part of our sensitivity analysis, we undertake firm-specific estimation and obtain results in correspondence with those reported in Tables 2 and 3. Although we just are not seeing evidence of widespread satisfaction-based mispricing under either approach, pursuing research frameworks that enhance power of mispricing tests would be advantageous.

3.4. How Widespread and What Is the Magnitude of Any Purported Customer Satisfaction-Based Mispricing?

In response to an earlier version of our manuscript, Fornell (2007) noted that for his top satisfaction portfolio, he too found very large abnormal returns for computer and Internet firms but stated that his findings of mispricing "are not limited to computer and Internet" firms. Unfortunately, we have not seen this evidence. The FMM2009 reported analyses neither (i) allowed for time-varying risk factor loadings nor (ii) presented disaggregate analysis capable of depicting how widespread any satisfaction-based mispricing might be. Based on their (or any) customer satisfaction-based portfolio formation criteria, do the financial markets misprice firms *not* in the computer and Internet sector? Allowing for time-varying risk covariates, what is the extent of this mispricing and is it statistically significant? Neither Fornell et al. (2006), FMM2009, nor Fornell et al. (2009b) report analyses that address these questions. We have no unique stake in the answers and would be interested in seeing evidence addressing these issues regardless of the particular outcomes.[2]

4. ILT2009: How Best to Assess the Pricing and Mispricing of Satisfaction?

We very much appreciate and recognize the conscientious and thoughtful effort of ILT2009 to provide

---

[2] Similarly, we would look forward to seeing further analyses of the effect of customer satisfaction on risk constructs. FMM2009 raise the possibility that customer satisfaction influences the differential between [upside beta – downside beta], that is, what they refer to as "beta surfing." Whereas FMM2009 are dismissive of work in finance on risk (e.g., p. 823), Low (2001) and Ang et al. (2006), for example, similarly advance the differential between upside beta and downside beta as being a relevant construct because evidence suggests it has a negative market risk premium. Ha (2009) assesses the extent to which brand attributes influence changes in this differential. Rigorous analyses of the effect of satisfaction on the upside–downside beta differential, including hypothesis tests with correctly specified type I errors, are warranted to see whether the FMM2009 conjectures can be substantiated. Furthermore, any enhanced model of expected return—for example, one encompassing postulated asymmetric effects of customer satisfaction on risk—could be used to generate estimates of mispricing. For example, a matched-firm approach taking into account downside risk, in addition to the standard matching characteristics of industry, size, and the book-to-market ratio, could be utilized. This analysis would assess the extent to which firms included in a top satisfaction portfolio have returns that differ from their matched counterpart.

an objective assessment of the initial and long-term financial market reaction to ACSI announcements. However, we disagree on some points regarding how these assessments should be carried out.

ILT2009 report three sets of analyses. The first two are not directly related to mispricing (i.e., the focus of the JM2009b study), but they warrant discussion because they have implications with respect to assessing the value relevance of customer satisfaction. The ILT2009 implementation of both of these analyses (i.e., assessing lagged effects of ACSI on accounting performance and event study analysis of annual ACSI announcements) has limitations. Their third analysis is directly related to mispricing and JM2009b. The reported findings on the lack of widespread mispricing are consistent with those we report as part of our sensitivity analyses, one of which involved (as does ILT2009) an assessment of mispricing based on a portfolio formation criteria of changes in satisfaction.

4.1. Assessing Lagged Effects of Customer Satisfaction on Accounting Performance Metrics vs. Direct Tests of Value Relevance

ILT2009's first set of analyses seeks to assess whether customer satisfaction leads accounting performance metrics. Their reported findings suggest that it does. However, as ILT2009 note, evidence of lagged effects of satisfaction on accounting performance measures does not indicate financial market mispricing. The financial markets may well realize the ACSI metric as a leading indicator of accounting performance measures. If so, market participants would react contemporaneously to unanticipated changes in the metric, and the metric would offer incremental explanatory power to accounting variables in explaining contemporaneous risk-adjusted stock returns. Jacobson and Mizik (2009a) undertake such direct tests of ACSI incremental value relevance.

A key limitation of assessing the impact on future operating performance measures for establishing potential incremental value relevance is that the findings are highly dependent on the estimating equations and the accounting metrics used. For example, the ILT2009 variable *Revenue* (which is constructed as revenue/assets) typically does not provide incremental information to ROA in explaining stock return. As such, it is unclear why its relationship with lagged ACSI offers insights incremental to that obtained from an analysis of the lagged relationship between ACSI and ROA.

ILT2009 (p. 826) argue that they make use of "well-established econometric methods." In fact, they do not make use of standard panel data econometric methods and tests, and as a result, their models have

questionable properties.[3] In particular, the ILT2009 accounting performance models do not allow for a fixed effects framework. To the extent that customer satisfaction is correlated with these time-invariant firm-specific fixed effects, the coefficient estimates on lagged customer satisfaction will be biased. Furthermore, their models also do not include industry-specific time-period dummy variables; i.e., they do not cross-sectionally demean the data.[4] ILT2009 clearly acknowledge the presence of firm-specific and time period-specific effects; i.e., they make use of cluster robust standard errors to adjust for the influence of these factors on standard errors. However, they assume, without testing, the lack of correlation of these effects with the explanatory variables in their models. This approach is at odds with standard econometric panel data methods and is directly counter to what recent econometric work in finance recommends (e.g., Petersen 2009, footnote 23). A good starting point for lead-lag analysis would be a panel data vector autoregressive (VAR) model (Holtz-Eakin et al. 1988)—that is, a more general specification that includes not only fixed effects but also higher-order lagged effects. A VAR approach encompasses a wider spectrum of possible models than what ILT2009 estimate.

Because ILT2009 do not report sensitivity analyses, the extent to which a more comprehensive model would affect conclusions, be it positively or negatively, is unclear. However, in Jacobson and Mizik

---

[3] We would be remiss in failing to note that the first two authors of ILT2009 have in the past made claims and advanced models of customer satisfaction that have not generalized well across subsequent analyses. For example, Ittner and Larcker (1998) undertake an event study where they conclude that the financial markets respond to the *level* of customer satisfaction, a finding that has had limited generalizablility and is counter to conclusions and methods they advance in ILT2009, which highlight the need to work with unanticipated changes in ACSI in the event study context. In Baum et al. (2000), they advance a levels model of market value and conclude: "Perhaps the most amazing result of our research is that two intangible asset categories—use of technology and customer satisfaction—had no statistical association with market values. That means these things, in contrast to our readers' perceptions, aren't helping companies create value at all. For all the blather over the past 10 years about the importance of customer satisfaction, it apparently has no effect on corporate value." We believe that some of the models advanced in ILT2009 may similarly lack the rigor needed to generate generalizable findings.

[4] Skoulakis (2006, p. 31) presents simulation results showing that "omitting the demeaning step can lead to a deterioration in the performance of the estimators, and inference based on *t*-statistics can be misleading." Furthermore, the consistency of cluster robust standard errors depends upon having a sufficient number of clusters. Simulation evidence (e.g., Petersen 2009) shows that cluster robust standard errors understate (albeit modestly) the true standard errors when the number of time-series observations is limited. This again highlights the usefulness of cross-sectionally demeaning the data or including industry-specific time-period dummy variables in the model.

(2009a), we did estimate a fixed effects VAR model for ROA (as a part of sensitivity analysis in developing an estimate of unanticipated ROA). Because the fixed effects model is a less efficient estimator, we first ran a Hausman specification test to assess the presence of fixed effects correlated with the explanatory factors. The test rejected the null hypothesis of no fixed effects; i.e., the estimated coefficient for lagged ACSI is biased in a model not allowing for a fixed effect and so warranted the use of the fixed effects models. In our fixed effects specification, lagged changes in customer satisfaction were not significantly related to the contemporaneous changes in ROA. These findings raise questions about the ILT2009 models (which do not make use of the fixed effects estimator) and conclusions related to the lagged effects of ACSI on operating performance.

Because of the various issues in modeling accounting series, we suggest that direct tests of incremental value relevance (such as those undertaken in Barth et al. 1998, Jacobson and Mizik 2009a, and other studies) are advantageous. These tests are based on the model of the form

$$\text{stkr}_{it} = \sum_{j=1}^{J} \beta_j * \tilde{\Delta} AccP_{it}$$
$$+ \gamma * \tilde{\Delta}\text{Non-Fin-Metric}_{it} + \varepsilon_{it}, \quad (\text{R1})$$

where stkr_{it} is the risk-adjusted return for firm i and time t, $\tilde{\Delta} AccP_{it}$ is the unanticipated change in accounting performance metric j for firm i and time t, and $\tilde{\Delta}\text{Non-Fin-Metric}_{it}$ is the unanticipated change in the non-financial performance metric for firm i and time t. A test of incremental value relevance involves an assessment of the statistical significance of γ.

In Jacobson and Mizik (2009a), we report that the ACSI measure has value relevance; that is, changes in ACSI have a statistically significant bivariate contemporaneous association with abnormal stock return. However, the metric appears limited in that it does not exhibit statistically significant value relevance *incremental* to accounting information. In other words, the effect of the change in satisfaction is no longer statistically significant in explaining risk-adjusted return once changes in accounting variables—in particular, changes in ROA—are also included in the model. The noise in the measure dominates any potential underlying signal in ACSI changes reflective of firm future profitability, which is incremental to information contained in accounting metrics. That is, changes in the ACSI measure contain too much noise/measurement error such that they are not providing statistically significant incremental value relevance to changes in current-term accounting information in explaining risk-adjusted return.

However, we should note again that these findings are not directly pertinent to mispricing analysis. The Corr(stkr$_{it}$, ΔX_{it}) (i.e., value relevance studies) will not be related to Corr(stkr$_{it+1}$, ΔX_{it}) or (stkr$_{it+1}$, X_{it}) (i.e., mispricing studies) as Corr(stkr$_{it}$, stkr$_{it+1}$) = 0 (i.e., risk-adjusted stock return is not autocorrelated). That is, studies of mispricing are orthogonal to studies of value relevance.

4.2. Event Study Analysis: Are the Annual Customer Satisfaction Announcements Timely Sources of Information to the Capital Markets?

ILT2009 undertake a second set of tests involving event study analyses. They find a short-term positive market response for the top quintile of ACSI firms (i.e., firms with the largest percent gains in ACSI score) but no association with abnormal return for the other quintiles. Here, we have few specific insights to share, as none of our analyses to date has focused on short-term market reactions. We can, however, offer some general thoughts.

First, because the firms included in quintiles change in every period, the risk factor loadings of quintiles can be expected to change as well. This consideration would suggest the need to allow for more general time-varying risk factor loadings.

Second, because only the top quintile shows abnormal return and no difference exists among the other quintiles, the question arises as to how widespread the short-term reaction is. Disaggregate analysis might be useful in helping to identify whether observed effects (or a lack of effects) are widespread or unique to select segments of firms. Particularly when a linear relationship is not found or when opposite effects between the top and bottom quintiles are not observed, the possibility exists that a finding may be unique to a select number of firms.

Third, and most important, measurement error in ILT2009's unanticipated ACSI estimate (stemming from a naïve model of expectations that does not allow market participants to use other information to adjust their expectations of the ASCI score) may well be a key driver responsible for failing to observe a widespread financial market reaction to changes in customer satisfaction. Although ILT2009 discuss the need to work with unanticipated components of ACSI in the event study setting (and we fully agree), the expectations model ILT2009 use implies market participants do not update their expectations of customer satisfaction throughout the year but only when the ACSI is announced. Because market participants have access to other sources of information likely to be correlated with the ACSI satisfaction metric (e.g., company announcements related to satisfaction, satisfaction surveys by other companies, Internet chatter,

customer complaints, quarterly financial results, etc.), expectations will be updated throughout the year so that the year-to-year change in satisfaction will not be a good measure of the unanticipated ACSI that exists at the time of the ACSI announcement. This induces measurement error in the estimate of the unanticipated change in satisfaction and generates a downward bias in estimated effects.

Kothari (2001, pp. 118–199) notes that in the context of earnings data, "competing sources of information (including quarterly earnings) preempt the information in annual earnings by about 85%. In this sense, annual accounting numbers are not a particularly timely source of information to the capital markets." Given the competing (i.e., correlated) sources of information available with respect to customer satisfaction, why should we expect annual customer satisfaction announcements to be particularly timely sources of information? Absent in-depth modeling of financial market expectations of customer satisfaction, this consideration preempts the usefulness of event study analysis for assessing the pricing of customer satisfaction.

A means of addressing the availability of correlated information about satisfaction, which is released on a more timely basis, is to extend the measurement window (Kothari 2001, p. 139) so as to allow for a financial market response throughout the year to changes in customer satisfaction information, which is what response modeling does. Financial market participants are able to impound into the stock price financial implications associated with customer satisfaction prior to the ACSI announcement as they gain this information from alternative sources. For example, investors do not need to wait for the ACSI announcement to know a dramatic event (such as the customer service announcements Nayyar 1995 studies) has affected customer satisfaction.

Because financial market participants likely gain information about satisfaction throughout the year, we prefer the use of the wider window response modeling (Equation (R1)) for assessing value relevance. Although ILT2009 seek to assess the pricing of customer satisfaction, their analyses ignore what might potentially be a more substantial effect, namely, the within-period contemporaneous response. This association can be assessed, as in Jacobson and Mizik (2009a), by expanding a standard earnings response model (e.g., Kormendi and Lipe 1987) to include non-financial measures (Equation (R1)). Aaker and Jacobson (1994) and Barth et al. (1998), for example, use this approach to assess the *incremental* information content of brand equity measures—that is, the extent to which brand equity metrics provide explanatory power incremental to accounting information in explaining stock return.

4.3. Mispricing Analysis

ILT2009 undertake a third study using the calendar-time portfolio approach for assessing mispricing that mirrors the approach Aksoy et al. (2008), O'Sullivan et al. (2009), JM2009b (Table 1, panel A), and FMM2009 (Table 1) use. It differs in that portfolios are formed based on the change in ACSI, and they have different holding periods for stocks in their portfolios. ILT2009 fail to find evidence of mispricing.

4.3.1. Portfolio Formation Criteria. ILT2009 contend that only changes in a variable should be used to assess mispricing and that this requirement (i) stems from efficient markets and (ii) is the approach accounting and finance advance. Financial market participants should only react to unanticipated changes in a variable in an efficient markets framework. As such, when assessing value relevance, unanticipated changes in a metric (rather than levels) should be linked to contemporaneous stock returns. However, once markets are allowed to be inefficient (acknowledging that mispricing might exist), no such restriction comes into play.

Indeed, the market may misvalue firms based on changes in a variable, levels of a variable, a combination of both levels and changes, or indeed, any transformation involving metrics in the available information set. Contrary to what ILT2009 state, the literature in accounting and finance is replete with studies that form portfolios based on levels rather than changes in variables. For example, as we mention in JM2009b, studies suggest firms may be mispriced based on the level of research and development (R&D) intensity (Chan et al. 2001) as well as unanticipated changes in R&D (Eberhart et al. 2004). Edmans (2008) reports that the firms included in "The 100 Best Companies to Work for in America" (i.e., a threshold level of employee satisfaction) might be mispriced. These studies are part of a literature stream suggesting financial markets may improperly value intangible assets. We do not have an explanation as to why ILT2009 overlook this literature stream (i.e., previous work forming portfolios based on autocorrelated levels variables to assess potential mispricing) so as to view work in marketing using similar formation criteria as "unusual."

No one right way and single formation criteria exist to form portfolios that can be assessed for mispricing. This fact adds another layer of difficultly to assessments of mispricing. Different portfolio formation criteria can lead to different conclusions regarding mispricing, which is why we assess a number of different satisfaction-based criteria and report that our conclusions are not affected. Although we examine a number of different portfolio formation criteria, we focus on one based jointly on levels and changes because (i) Fornell et al. (2006, p. 8) state that it

seems reasonable to base a trading strategy on both levels and changes, and (ii) Aksoy et al. (2008) report statistically significant findings using this combined portfolio formation rule. Although results pertaining to mispricing are not invariant to the specific manner in which portfolios are formed, we fail to find any satisfaction-based portfolio formation rule that produces evidence of widespread mispricing.

4.3.2. Time-Varying Risk Factor Loading. The mispricing analyses in ILT2009 do not allow for time-varying risk factor loading even though the composition of firms in the portfolios changes over time. The role of risk plays a larger role in long-term mispricing studies than in short-term event studies; thus, the model for expected return becomes more central. As such, assessing the sensitivity of results to models of expected return would be prudent. The inappropriateness of assuming constant risk factor loadings when firms in a portfolio change over time is widely acknowledged. In addition to the references we cited previously, Shivakumar (2000, p. 365), for example, is unambiguous in stating that the assumption of constant risk factor loadings "is untenable given that the composition and number of firms in the quintiles change constantly." This condition is clearly applicable to the ILT2009 portfolios, yet they assume constant risk factor loadings. Particularly when the composition of the portfolio is not constant over time or the time period for assessment is lengthy, using time-varying risk factor loadings should be part of future research assessing potential mispricing.

5. Concluding Remarks

Perhaps it goes without saying, but we do not take issue with, neither are we surprised by, ILT2009's failure to find evidence of mispricing even for computer and Internet firms. We are well aware that conclusions concerning mispricing may be sensitive to, for example, the model for expected return, the time period for analyses, the sample of firms included in the study, and specific criteria used to form portfolios.

Although studying mispricing is certainly of value, we have noted previously (Mizik and Jacobson 2009) that studies of this type may be of limited usefulness if the goal is to establish empirical regularities or generalizations. Market participants have the incentives to and thus do dissipate financial market anomalies. Reported anomalies may be highly time- and context-specific and may disappear or even switch signs (Shiller 2002).

Failure to find mispricing says little about the value implications of a metric; it says more about financial-market efficiency in incorporating information. A lack of statistically significant evidence of mispricing is simply consistent with financial markets being efficient and the markets reacting (or the not reacting) contemporaneously to information (or lack of valuable information) contained in a measure so that no lagged relationship between a metric and future-term abnormal stock return exists.

Value-relevance studies are likely to have greater applicability to marketing applications. Do changes in a measure provide incremental information to changes in accounting metrics in explaining contemporaneous risk-adjusted stock returns? The Jacobson and Mizik (2009a) failure to find statistically significant incremental value relevance for the ACSI metric is far more disappointing to those advancing the ACSI metric as a supplement to financial performance than the lack of evidence supportive of mispricing. The findings together highlight the need for additional theoretical, empirical, and measurement work to better understand the workings of customer satisfaction.

References

Aaker, D. A., R. Jacobson. 1994. The financial information content of perceived quality. *J. Marketing Res.* **31**(2) 191–201.

Aksoy, L., B. Cooil, C. Groening, T. L. Keiningham, A. Yalçin. 2008. The long-term stock market valuation of customer satisfaction. *J. Marketing* **72**(4) 105–122.

Anderson, E. W., C. Fornell, R. T. Rust. 1997. Customer satisfaction, productivity, and profitability: Differences between goods and services. *Marketing Sci.* **16**(2) 129–145.

Ang, A., D. Kristensen. 2009. Testing conditional factor models. Working paper, Columbia University, New York.

Ang, A., J. Chen, Y. Xing. 2006. Downside risk. *Rev. Financial Stud.* **19**(4) 1191–1239.

Barth, M. E., M. B. Clement, G. Foster, R. Kasznik. 1998. Brand values and capital market valuation. *Rev. Accounting Stud.* **3**(1–2) 41–68.

Baum, G., C. Ittner, D. Larcker, J. Low, T. Siesfeld, M. S. Malone. 2000. Introducing the new Value Creation Index. *Forbes ASAP* (April 3), http://www.forbes.com/asap/2000/0403/140.html.

Biørn, E. 2001. The efficiency of panel data estimators: GLS versus estimators which do not depend on variance components. Working Paper 28/2001, Department of Economics, University of Oslo, Oslo, Norway.

Chan, L. K. C., J. Lakonishok, T. Sougiannis. 2001. The stock market valuation of research and development expenditures. *J. Finance* **56**(6) 2431–2456.

Chow, G. C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28**(3) 591–605.

Eberhart, A. C., W. F. Maxwell, A. R. Siddique. 2004. An examination of long-term abnormal stock returns and operating performance following R&D increases. *J. Finance* **59**(2) 623–650.

Edmans, A. 2009. Does the stock market fully value intangibles? Employee satisfaction and equity prices. Working paper, The Wharton School of the University of Pennsylvania, Philadephia. http://ssrn.com/abstract=985735.

Fama, E. F. 1998. Market efficiency, long-term returns, and behavioral finance. *J. Financial Econom.* **49**(3) 283–306.

Fama, E. F., K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* **33**(1) 3–56.

Fama, E. F., K. R. French. 1997. Industry costs of equity. *J. Financial Econom.* **43**(2) 153–193.

Fornell, C. 2007. Personal communication, e-mail. (April 5).

Fornell, C., S. Mithas, F. V. Morgeson III. 2009a. The economic and statistical significance of stock returns on customer satisfaction. *Marketing Sci.* **28**(5) 820–825.

Fornell, C., S. Mithas, F. V. Morgeson III. 2009b. The statistical significance of portfolio returns. *Internat. J. Res. Marketing* **26**(2) 162–163.

Fornell, C., S. Mithas, F. V. Morgeson III, M. S. Krishnan. 2006. Customer satisfaction and stock prices: High returns, low risk. *J. Marketing* **70**(1) 3–14.

Ha, K. N. 2009. Asymmetric risk implications of brand equity dimensions. Working paper, Foster School of Business, University of Washington, Seattle.

Holtz-Eakin, D., W. Newey, H. S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* **56**(6) 1371–1395.

Hsiao, C. 1986. *Analysis of Panel Data*. Cambridge University Press, Cambridge, UK.

Ittner, C., D. Larcker. 1998. Are non-financial measures leading indicators of financial performance? An analysis of customer satisfaction. *J. Accounting Res.* **36**(Supplement) 1–35.

Ittner, C., D. Larcker, D. Taylor. 2009. The stock market's pricing of customer satisfaction. *Marketing Sci.* **28**(5) 826–835.

Jacobson, R., N. Mizik. 2009a. Assessing the value-relevance of customer satisfaction. Working paper, http://ssrn.com/abstract=990783.

Jacobson, R., N. Mizik. 2009b. The financial markets and customer satisfaction: Reexamining possible financial market mispricing of customer satisfaction. *Marketing Sci.* **28**(5) 810–819.

Kormendi, R., R. Lipe. 1987. Earnings innovations, earnings persistence and stock return. *J. Bus.* **60**(3) 207–38.

Kothari, S. P. 2001. Capital markets research in accounting. *J. Accounting Econom.* **31**(1–3) 105–231.

Kothari, S. P., J. B. Warner. 2001. Evaluating mutual fund performance. *J. Finance* **56**(5) 1985–2010.

Kothari, S. P., J. B. Warner. 2007. Econometrics of event studies. B. E. Eckbo, ed. *Handbook of Corporate Finance*, Vol. 1. North-Holland, Amsterdam, 3–36.

LeRoy, S. F. 1989. Efficient capital markets and martingales. *J. Econom. Literature* **27**(4) 1583–1621.

Lewellen, J., S. Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *J. Financial Econom.* **82**(2) 289–314.

Low, C. 2001. Semidimensional risks in the cross section of stock returns. Working paper, National University of Singapore, Singapore.

Mizik, N., R. Jacobson. 2009. Financial markets research in marketing. *J. Marketing Res.* **46**(3) 320–324.

Nayyar, P. R. 1995. Stock market reactions to customer service changes. *Strategic Management J.* **16**(1) 39–53.

O'Sullivan, D., M. C. Hutchinson, V. O'Connell. 2009. Empirical evidence of the stock market's (mis)pricing of customer satisfaction. *Internat. J. Marketing Res.* **26**(2) 154–161.

Pástor, L., P. Veronesi. 2007. Technological revolutions and stock prices. Working paper, Graduate School of Business, University of Chicago, Chicago.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Rev. Financial Stud.* **22**(1) 435–480.

Shiller, R. J. 2002. From efficient market theory to behavioral finance. Discussion Paper 1385, Cowles Foundation for Research in Economics, Yale University, New Haven, CT.

Shivakumar, L. 2000. Do firms mislead investors by overstating earnings before seasoned equity offerings? *J. Accounting Econom.* **29**(3) 339–371.

Skoulakis, G. 2006. Panel data inference in finance: Least-squares vs. Fama-MacBeth. Working paper, University of Maryland, College Park.

Taylor, W. E. 1980. Small sample considerations in the estimation from panel data. *J. Econometrics* **13**(2) 203–223.