

Bell, David R.; Ledoit, Olivier; Wolf, Michael

Working Paper

Reexamining possible mispricing of customer satisfaction

Working Paper, No. 79

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Bell, David R.; Ledoit, Olivier; Wolf, Michael (2012) : Reexamining possible mispricing of customer satisfaction, Working Paper, No. 79, University of Zurich, Department of Economics, Zurich,
<https://doi.org/10.5167/uzh-62713>

This Version is available at:

<https://hdl.handle.net/10419/77553>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 79

Reexamining Possible Mispricing of Customer Satisfaction

David R. Bell, Olivier Lledoit and Michael Wolf

May 2012

Reexamining Possible Mispricing of Customer Satisfaction

David R. Bell
Marketing Department
The Wharton School
Philadelphia, CA 19104-6340, USA
davidb@wharton.upenn.edu

Olivier Ledit
Department of Economics
University of Zurich
CH-8032 Zurich, Switzerland
olivier.ledoit@econ.uzh.ch

Michael Wolf*
Department Economics
University of Zurich
CH-8032 Zurich, Switzerland
michael.wolf@econ.uzh.ch

May 2012

Abstract

There has been a recent debate in the marketing literature concerning the possible mispricing of customer satisfaction. While earlier studies claim that portfolios with attractive out-of-sample properties can be formed by loading on stocks whose firms enjoy high customer satisfaction, later studies challenge this finding. A large part of the disagreement stems from the difficulty of how to actually evaluate mispricing based on the observed portfolio returns. In particular, any portfolio formation method that requires the use of a risk model is open to the criticism of time-varying risk factor loadings due to the changing composition of the portfolio over time. As an alternative, we construct portfolios that are neutral with respect to the desired risk factors *a priori*. Consequently, no risk model is needed when analyzing the observed returns of our portfolios. Using various ways of measuring customer satisfaction, we do not find any convincing evidence that portfolios that load on high customer satisfaction lead to abnormal returns.

KEY WORDS: customer satisfaction, financial performance, long-short portfolio, mispricing.

JEL CLASSIFICATION NOS: G11, G12.

*Research has been supported by the NCCR Finrisk project “New Methods in Theoretical and Empirical Asset Pricing”.

1 Introduction

Most managers and marketing academics would like to believe that marketing actions can contribute to financial value. Firms with strong brands and satisfied customers might therefore be expected to outperform their competitors and the market.¹ While academics have examined the financial performance consequences of reputation, brand image, and product quality, work on customer satisfaction has been the most widespread (and the most controversial). The controversy stems not from disagreements about whether “customer satisfaction” is a worthwhile management objective, but rather whether or not financial markets accurately account for it. A review article by [Mittal and Frennea](#) (2010, p.2) is unequivocal on the former point:

Two decades of academic research have quantified the impact of customer satisfaction on a number of beneficial customer behaviors and consequent financial performance. It is clear that firms that manage their customers as well as costs realize greater financial returns compared to firms that ignore customer satisfaction.

Yet, starting with [Fornell et al. \(2006\)](#), there have been several academic studies concerning whether the (potential) mispricing of customer satisfaction can be exploited to form investment portfolios with abnormal returns. This point is again nicely made by [Mittal and Frennea](#) (2010, p.4) in their review:

There has been recent discussion about whether customer satisfaction can predict abnormal stock-market returns, although additional data and theoretical development are needed to resolve this issue.

The question of abnormal returns, and in particular, how to properly investigate them, is the focus of this paper.

The basic approach taken by previous papers that have examined mispricing is (i) to decide on a specific portfolio formation rule; (ii) to use this rule on past data in order to observe corresponding returns; and (iii) to examine whether the observed returns are ‘abnormally’ good. In particular, it seems agreed that positive findings in (iii) should be backed up by tests of statistical significance. There are two major ways in which such portfolios can be formed. First, one can form long-only portfolios by simply buying stocks of firms with high customer satisfaction. Second, one can form long-short portfolios by buying stocks of firms with high customer satisfaction while, at the same time, selling stocks of firms with low customer satisfaction. (Note that a mutual fund would only be allowed to use long-only portfolios whereas a hedge fund would also be allowed to use long-short portfolios.) After deciding on

¹The *Marketing Science Institute*, for example, has for many years considered research on the financial value of marketing a top priority for research.

a particular portfolio formation rule and observing the resulting returns, one is then left to judge whether those returns are indeed ‘abnormally’ good.

For long-only portfolios, this question necessitates the use of an appropriate risk model. The reason is that by simply holding a portfolio of stocks, one is ‘guaranteed’ a positive expected return already. A risk model takes into account the extent to which a portfolio is exposed to standard risk factors, such as the market, size, book-to-market, and momentum. It is then fair to examine whether the ‘left-over’ returns, after adjusting for risk factors, have a positive expected value. The problem with this approach is that it assumes the exposures to the various risk factors to be constant over time. Unfortunately, this assumption is hardly fulfilled for portfolios formed on customer satisfaction, since the portfolio composition generally changes over time. As a result, studies adopting this approach are always open to a certain amount of criticism.

For long-short portfolios, the situation seems different at first. Since some stocks are bought while others are sold, it would seem fair to simply test for an overall positive expected portfolio return. But even long-short portfolios can have exposure to risk factors. For example, this can happen if the firms in the long portfolio have larger book-to-market, on average, than the firms in the short portfolio. So in the end, in general, a risk model is needed again to establish whether the observed portfolio returns are ‘abnormally’ good. And the same criticism concerning non-constant exposures to risk factors that held for the long-only portfolio holds here as well.

In this paper, we develop and adopt a novel approach that eliminates the need to use a risk model altogether. By ‘novel’, we mean this with respect to the particular strand of literature examining mispricing of customer satisfaction. What we do is to look at the problem through the eyes of a real-world fund manager aiming to form a long-short portfolio based on customer satisfaction with zero exposure *a priori* to any risk factors considered. Hence, the constraint of no exposure to risk factors must be built into the portfolio formation rule at the outset. If this is done, one can indeed simply test for a positive expected portfolio return, as there is no longer a need for an adjustment for risk factors *a posteriori*. While we apply our method to the mispricing of customer satisfaction, it could also be used to assess the impact of other marketing factors. Thus, our methodological contribution is quite general and should be of interest to other researchers who measure financial performance consequences of various marketing actions, including investments in brand, product quality, and so so. Our substantive contribution is the following: after analyzing a wide range of scenarios we do not find any convincing evidence for the mispricing of customer satisfaction. This does not imply that investments in satisfaction, *per se*, are not worthwhile; rather, that the cross-sectional discrepancies in satisfaction scores and their temporal evolution are fully accounted for by the

market.

The remainder of the paper is organized as follows. Section 2 reviews previous findings from the literature on possible mispricing of customer satisfaction. Section 3 provides the details of our new approach for portfolio formation and reports corresponding results after applying such rules to real data. Finally, Section 4 concludes the paper with a brief discussion and implications for future research.

2 Background and Review of Previous Findings

This section reviews some recent findings in the literature that serve to motivate our own study.² The specific strand of literature dates back to Fornell et al. (2006) who argue that portfolios that load on stocks of firms enjoying high customer satisfaction outperform regular stock indexes such as DJIA, S&P 500, and NASDAQ. Apparently, these findings have not been considered definitive by researchers in the field for two main reasons. First, some of the portfolio formation criteria appear a bit arbitrary and have been suspected as potential in-hindsight maximizers; for example, see Ittner et al. (2009). Second, the findings are primarily descriptive in nature; ideally, outperformance should be backed up by demonstrating statistical significance. Subsequent studies, therefore, tend to use more clear-cut portfolio formation rules and also employ statistical tests of significance in assessing portfolio outperformance.

2.1 Long-Only Portfolios

Our point of reference is the study of Jacobson and Mizik (2009b). Their first contribution is to try and replicate a previous study of Aksoy et al. (2008), constructing long-only portfolios based on both level and recent changes of customer satisfaction. Firm-level customer satisfaction scores (on a 0–100 scale) can be downloaded free of charge from the American Customer Satisfaction Index (ACSI) database available at <http://www.theacsi.org>. ACSI collects and releases its data on an annual basis, but does so throughout the year in different waves for firms in different industries. Until June 2010 this was done according to a quarterly schedule; since then a monthly schedule has been in use.

Jacobson and Mizik (2009b) group firms into four portfolios based on whether (1) the firm's customer satisfaction score was above or below the national average for that time period and (2) the firm's customer satisfaction score was increasing or decreasing with respect to the previous year. The four resulting portfolios are as follows:

²A large literature considers a number of financial consequences resulting from customer satisfaction, including financial risk, e.g., Tuli and Bharadwaj (2009), analyst recommendations, e.g., Luo et al. (2010), and related issues. We focus exclusively on mispricing and therefore limit our review to relevant articles in this stream of work.

	Level of ACSI score compared to national average	Change in ACSI
Portfolio 1	Greater	Positive
Portfolio 2	Lower	Positive
Portfolio 3	Greater	Negative
Portfolio 4	Lower	Negative

It is expected that Portfolio 1 performs the best while Portfolio 4 performs the worst, with Portfolios 2 and 3 constituting an intermediate ‘gray area’; see [Aksoy et al. \(2008\)](#).

Since all four portfolios are long-only, mispricing cannot be tested by checking whether the average return is significantly different from zero. Instead, one needs to focus on the intercept in a suitable risk model. [Jacobson and Mizik \(2009b\)](#) employ the standard [Carhart \(1997\)](#) four-factor risk model:

$$\text{Ret}_{p,t} - \text{Ret}_{\text{rf},t} = \alpha_p + \beta_p \text{MKT}_t + \gamma_p \text{SMB}_t + \delta_p \text{HML}_t + \kappa_p \text{MOM}_t + e_{p,t} . \quad (1)$$

Here, $\text{Ret}_{p,t}$ denotes the return of portfolio p during month t ; $\text{Ret}_{\text{rf},t}$ denotes the return of the risk-free rate during month t ; MKT_t denotes the return of the market during month t in excess of the risk-free rate; and SMB_t , HML_t , and MOM_t denote the returns of the remaining three risk factors during month t : market size (SMB), book-to-market (HML), and momentum (MOM).

A portfolio is then considered *mispriced* if $\alpha_p \neq 0$. In particular, a portfolio with a positive (negative) intercept α_p is considered to deliver abnormally high (low) returns adjusted for systematic risks. Proponents of the theory of mispricing of customer satisfaction argue that α_1 should be positive and α_4 should be negative. There are no clear postulates concerning α_2 or α_3 , since these two portfolios each exhibit ‘mixed signals’.

Both [Aksoy et al. \(2008\)](#) and [Jacobson and Mizik \(2009b\)](#) start the portfolio formation process in Q3/1996 and end in Q1/2006 and therefore have $T = 117$ months of out-of-sample returns. Portfolios are rebalanced each time after new ACSI data are released and this rebalancing always occurs at the beginning of the third month of the quarter.³ The main focus of interest in these studies is the significance of $\hat{\alpha}_1$ and both studies find a point estimate of around 0.005. While [Aksoy et al. \(2008\)](#) obtain a t -statistic of 2.49 [Jacobson and Mizik \(2009b\)](#) find a smaller value of only 1.82. They speculate that the difference may be due to a slightly different universe of firms used in the two respective studies. While this is possible, another reason might lie in how the two respective standard errors of $\hat{\alpha}_1$ were computed. [Aksoy et al. \(2008\)](#) use

³New ACSI data were released during the months of February, May, August, and November until June 2010.

the heteroskedasticity and autocorrelation (HAC) standard errors of [Newey and West \(1987\)](#), without specifying their choice of bandwidth, while [Jacobson and Mizik \(2009b\)](#) do not say.⁴

The t -statistic of [Jacobson and Mizik \(2009b\)](#) is significant at the 5% level if a one-sided test is carried out, but only at the 10% level if a two-sided test is carried out. [Jacobson and Mizik \(2009b\)](#) use a two-sided test while [Fornell et al. \(2009\)](#) argue that a one-sided test should be used instead due to previous findings in the literature concerning desirable properties of portfolios that load on high customer satisfaction. Our point of view is that, indeed, a two-sided test is preferred, as the *a priori* view on the sign of α_1 is based on previous empirical findings (using largely overlapping time periods) rather than on economic theory. None of the other $\hat{\alpha}_p$ estimates is established as significant, though $\hat{\alpha}_4$ does have a negative sign in both studies.

Further evidence is provided by [O’Sullivan et al. \(2009\)](#) who re-examine the specific trading strategy of [Fornell et al. \(2006\)](#) but subject it to statistical tests based on the four-factor risk model (1). The trading strategy ranks stocks (firms) according to their ACSI scores and then groups stocks into quintiles accordingly. [Fornell et al. \(2006\)](#) find very attractive properties of the top-quintile portfolio, but only provide descriptive measures. [O’Sullivan et al. \(2009\)](#), on the other hand, fail to find statistical significance. They consider two investment periods: 02/1997 until 05/2003, which corresponds to the investment period of [Fornell et al. \(2006\)](#), and 03/1996 until 05/2006, which corresponds to the investment period of [Jacobson and Mizik \(2009b\)](#). The two resulting t -statistics for $\hat{\alpha}_p$ in (1) are 0.73 and 0.84, respectively.⁵ Finally, [O’Sullivan et al. \(2009\)](#) also examine a trading strategy closely related to the one of [Jacobson and Mizik \(2009b\)](#). The t -statistic for $\hat{\alpha}_p$, for the sample period 03/1996 until 05/2006, is 0.93.

A problem with all the studies above, as discussed by [Jacobson and Mizik \(2009b\)](#), is that the composition of any portfolio formed on customer satisfaction will change over time, as some firms will ‘move’ from one portfolio to another, while others ‘enter’ the universe at some intermediate point and yet others ‘disappear’ from the universe at some intermediate point. Hence, it is unrealistic to assume that the coefficients α_p , β_p , γ_p , δ_p , and κ_p remain constant over time. The failure of this key assumption thus renders the all the above findings questionable to some extent. It is important to note that, in principle, the failure of the assumption can bias t -statistics upwards or downwards, depending on circumstances. Thus,

⁴As an aside, we are somewhat suprised at the popularity that the sub-optimal [Newey and West \(1987\)](#) HAC standard errors continue to enjoy with empirical researchers in the areas of economics, finance, and marketing. HAC standard errors with better properties have been around for a very long time and were introduced to the social sciences by [Andrews \(1991\)](#) and [Andrews and Monahan \(1992\)](#) at the very latest.

⁵As alternatives to the four-factor risk model, [O’Sullivan et al. \(2009\)](#) also employ the market model and the three-factor risk model of [Fama and French \(1993\)](#). The resulting t -statistics for $\hat{\alpha}_p$ are then even smaller in these cases.

there is no straightforward consequence of this model-based problem for the answer to the substantive question of interest.

As a potential remedy, [Jacobson and Mizik \(2009b\)](#) also consider an alternative approach that allows for time-varying risk factors β_p , γ_p , δ_p , and κ_p , using a methodology proposed by [Lewellen and Nagel \(2006\)](#). After employing this procedure they find a t -statistic of 1.18 for $\hat{\alpha}_1$. While this new approach overcomes some of the shortcomings of the predominant portfolio-based approaches, it is not completely free of problems either. First, one needs to use daily return data for the firms in the portfolio universe in order to estimate the time varying risk-factors using a rolling window method. Second, both current and lagged risk factors have to be included in the daily extension of risk model (1). Third, and most important, it is still assumed that α_p does not change over time; otherwise, a test on ‘the’ intercept could not even be considered. But such an assumption appears highly questionable—if the risk factors are allowed to change over time, why can the intercept taken to be constant? This same criticism also applies to a second approach of [Jacobson and Mizik \(2009b\)](#); specifically, to allow for time-varying risk factors based on the methodology proposed by [McAlister et al. \(2007\)](#).

We therefore conclude that studying the potential mispricing of customer satisfaction using long-only portfolios will always be somewhat controversial. On the one hand, it can be argued that risk models with constant (over time) risk factors are inappropriate. On the other hand, if one allows for time-varying risk factors, then, arguably, one should also allow for a time-varying intercept. But then it is not entirely clear anymore how a ‘clean’ test for mispricing can be carried out, let alone formulated.⁶

2.2 Long-Short Portfolios

Since long-only portfolios are clearly controversial, it seems natural to employ zero-investment long-short portfolios instead (or at least in addition). The motivation is that a long-short portfolio is considered successful if, simply, it delivers a positive expected return. [Aksoy et al. \(2008\)](#) construct a long-short portfolio P1 – P4; that is, in each investment period, they go long one unit Portfolio 1 and short one unit Portfolio 4.⁷ For this portfolio, they find an average out-of-sample return of 0.0092 per month with an associated t -statistic of 2.30.

While this long-short approach has conceptual merit, it is open to the criticism that even a long-short portfolio might unwittingly load on a well-known risk factor, such as beta or book-to-market. This happens, for example, if the long portfolio contains firms with a larger beta,

⁶One possibility would be to consider a time-varying intercept given by a ‘base’ intercept plus mean-zero, period-specific deviations. The test would then concern the ‘base’ intercept. Apparently, such an approach was also tried by [Jacobson and Mizik \(2009b\)](#) but not included in the paper, as it gave results that were very similar to those from the models that used a fixed intercept. This was communicated to us by [Mizik \(2011\)](#).

⁷They call the resulting portfolio “High – Low” instead.

on average, than the firms in the short portfolio. As a result, a significant positive average return for the long-short portfolio could, in principle, be attributed to risk factor loadings as opposed to mispricing of customer satisfaction.

To address this concern, one can again apply a risk model to the returns of the long-short portfolio, such as the four-factor model of [Carhart \(1997\)](#):

$$\text{Ret}_{\text{LS},t} = \alpha_{\text{LS}} + \beta_{\text{LS}} \text{MKT}_t + \gamma_{\text{LS}} \text{SMB}_t + \delta_{\text{LS}} \text{HML}_t + \kappa_{\text{LS}} \text{MOM}_t + e_{\text{LS},t} . \quad (2)$$

Here, $\text{Ret}_{\text{LS},t}$ denotes the return of the long-short portfolio in period t and the right hand side regressors are defined as in (1). Using this formulation, [Aksoy et al. \(2008\)](#) find a point estimate of $\hat{\alpha}_{\text{LS}} = 0.0088$ with an associated t -statistic of 2.22. By reasoning analogous to that given previously, this empirical finding is also open to the criticism of time-varying risk factor loadings and, arguably, time-varying intercept as well in (2).

An alternative long-short portfolio is formed by [Ittner et al. \(2009\)](#). Instead of using both the levels and the recent changes of ACSI scores, they only use the recent changes. Furthermore, instead of simply using the sign (that is, up versus down movement), they use the actual change expressed as a percentage. [Ittner et al. \(2009\)](#) argue that using changes but not levels corresponds to “the more typical accounting and finance practice of measuring the amount of new (or unexpected) information provided to the market”, at least for a variable that is autocorrelated over time such as the ACSI score. The argument for using the percentage change rather than the sign only is that it contains more information. We agree with the latter argument that the percentage change contains more information but are less sure about the former (that the finance literature eschews variables in levels). The finance literature contains many examples of variables, which, for asset pricing purposes, are used in levels as opposed to in recent changes even though the variables are autocorrelated over time, such as market size and book-to-market. Concerning this question we agree with [Jacobson and Mizik \(2009a\)](#) who state that “No one right way and single formation criteria exist to form portfolios that can be assessed for mispricing”.

[Ittner et al. \(2009\)](#) group firms into quintiles depending on the percentage change of their respective ACSI score. They then go long the firms in the highest quintile and go short the firms in the lowest quintile. The resulting portfolio is called “Q5 – Q1”. Updating occurs every quarter after new ACSI scores are released. [Ittner et al. \(2009\)](#) use daily return data and consider holding periods of 365 days, 180 days, and one quarter (that is, until the next wave of ACSI scores is released). In total, they compute six t -statistics: there are three holding periods and two parameters of interest (the raw expected return and the intercept of risk model).⁸ All six t -statistics lie between 0.72 and 0.83. Obviously, these values are in

⁸[Ittner et al. \(2009\)](#) employ the three-factor model of [Fama and French \(1993\)](#).

stark contrast to the findings of Aksoy et al. (2008). While Ittner et al. (2009) use a bigger universe of firms (243) compared to Aksoy et al. (2008), who use 151 firms and a slightly longer investment period⁹, these differences hardly seem capable of explaining the large discrepancies in t -statistics. One is therefore left to speculate whether the way Ittner et al. (2009) use the ACSI scores in portfolio formation (that is, using only the percentage change) is somehow less informative compared to the way Aksoy et al. (2008) use them (that is, using both levels and the signs of the changes). We will return to this issue in our empirical study below.

3 A New Portfolio Formation Approach

Rather than looking at the problem of potential mispricing of customer satisfaction solely from the viewpoint of an academic researcher, we will put ourselves in the shoes of a hedge fund manager in the ‘real world’. Posing the question bluntly, we ask whether the potential mispricing can be used to actually make money with a long-short portfolio of the sort that would be employed by a ‘real-world’ fund manager. In particular, we want to form portfolios that *do not* load on any risk factors in the four-factor model (2) of Carhart (1997). In other words, we aim to be market-neutral (that is, β_{LS} equal to zero), size-neutral (that is, γ_{LS} equal to zero), book-to-market neutral (that is, δ_{LS} equal to zero), and momentum-neutral (that is, κ_{LS} equal to zero) from the outset.

The key advantage is the following: Under this approach, the question of mispricing can be answered by simply looking the expected return of the resulting portfolio, as this is the quantity that corresponds one-to-one to α_{LS} in (2) *if* all four risk factors loadings are equal to zero. In other words, by properly accounting for the four risk factors *a priori* (that is, at the portfolio formation stage), we no longer need a risk model in the evaluation of the portfolio returns *a posteriori*. The beauty of this approach is that it thereby completely eliminates the issue of time-varying coefficients. Note here that a long-short portfolio is the only way to achieve neutrality with respect to all four risk factors when investing in stocks of large firms such as those as covered by the ACSI database. Obviously any long-only portfolio, for example, will have some exposure to the market and therefore cannot be beta-neutral.

3.1 Achieving Risk-Factor Neutrality

The question now becomes how to make sure that our portfolios will have zero exposure to the four risk factors from the outset and by construction. The answer is that the portfolio weights must be chosen in a way such that all four constraints on the coefficients are necessarily and

⁹They use ACSI scores from Q1/1995 until Q4/2006 but do not exactly specify the out-of-sample period for which they observe the returns of their portfolios.

simultaneously satisfied. Assume that, at a given point in time, we have a universe of N firms from which we want to form a long-short portfolio. Denote the weight of company i in the portfolio by w_i , for $i = 1, \dots, N$. A long-short portfolio satisfies

$$\sum_{i=1}^N w_i = 0 \quad \text{with} \quad \sum_{i=1}^N |w_i| > 0, \quad (3)$$

where the latter condition rules out an ‘empty’ portfolio.

As further ingredients, we need to measure beta, market size, book-to-market, and momentum for each company. Denote the corresponding quantities, for company i , by beta_i , size_i , btm_i , and mom_i , respectively. Needless to say, all of these quantities must be measured with information that is available on the day that the portfolio is formed. An important consideration here is that the distribution of the quantity size_i tends to be very much skewed to the right in just about any collection of firms. As a remedy, it is standard to take the logarithm of market size, which results in a more bell-shaped distribution. Denote this quantity by log-size_i for company i . The four neutrality constraints are then expressed as

$$\sum_{i=1}^N w_i \cdot \text{beta}_i = 0, \quad \sum_{i=1}^N w_i \cdot \text{log-size}_i = 0, \quad \sum_{i=1}^N w_i \cdot \text{btm}_i = 0, \quad \text{and} \quad \sum_{i=1}^N w_i \cdot \text{mom}_i = 0. \quad (4)$$

3.2 Loading on Customer Satisfaction

Any portfolio satisfying (3)–(4) would be a valid choice for our purposes in the sense of being a portfolio that is long-short and neutral with respect to the four risk factors. But so far no information on customer satisfaction has been built in. Let cs_i denote a specific measure of customer satisfaction for company i . We then aim, over all ‘valid’ portfolios, to maximize the satisfaction measure of the portfolio expressed as

$$\sum_{i=1}^N w_i \cdot \text{cs}_i. \quad (5)$$

Needless to say, the choice of customer satisfaction measure is not unique. One possibility, in the spirit of [Aksoy et al. \(2008\)](#) and [Jacobson and Mizik \(2009b\)](#), is to take the sum of two indicator functions: whether the ACSI score of company i is above the national average and whether the last change has been positive:

$$\text{cs}_i = \mathbb{1}\{\text{score}_i > \text{national average}\} + \mathbb{1}\{\text{score}_i > \text{previous score}_i\}. \quad (6)$$

Following this approach, we get $\text{cs}_i=0$ for firms in their Portfolio 4, $\text{cs}_i=1$ for firms in their Portfolios 2 and 3, and $\text{cs}_i=2$ for firms in their Portfolio 1. On the other hand, the proposal of [Ittner et al. \(2009\)](#) corresponds to

$$\text{cs}_i = \frac{\text{score}_i}{\text{previous score}_i} - 1. \quad (7)$$

At this point, we introduce a more general measure of customer satisfaction that includes (7) as a special case but, potentially, also incorporates the levels. A naïve approach would be to simply take a weighted average of levels and percentage changes but since these two quantities live on different scales, this would be akin to combining apples with oranges. Before taking a weighed average, one has to make sure that the two scales are comparable. This can be achieved by the standard approach of converting the two quantities to z -scores, that is, by first subtracting the cross-sectional average and then dividing by the cross-sectional standard deviation.

For a general collection of numbers a_1, a_2, \dots, a_N , denote

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i \quad \text{and} \quad s_a^2 = \frac{1}{N} \sum_{i=1}^N (a_i - \bar{a})^2 . \quad (8)$$

Then the z -score corresponding to a_i is defined as

$$z(a_i) = \frac{a_i - \bar{a}}{s_a} . \quad (9)$$

To keep the notation compact we introduce the following two definitions:

$$z_{l,i} = z(\text{score}_i) \quad \text{and} \quad z_{c,i} = z(\text{score}_i / \text{previous score}_i - 1) , \quad (10)$$

where the subscript l stands for ‘level’ and the subscript c stands for (percentage) ‘change’. As a more general measure of customer satisfaction, we then propose

$$cs_i = \rho \cdot z_{l,i} + (1 - \rho) \cdot z_{c,i} , \quad \text{with } \rho \in [0, 1] . \quad (11)$$

This includes (7) as a special case when choosing $\rho = 0$. On the other end of the spectrum, choosing $\rho = 1$ only uses the levels. A compromise using both levels and percentage changes, with equal weights, can be obtained by choosing $\rho = 0.5$.

3.3 Portfolio Formation in Full

The portfolio formation objective, so far, is to maximize the satisfaction measure (5) subject to the constraints (3)–(4). That is, to form a “maximal satisfaction” portfolio of stocks which *a priori* does not load on any risk factors in the four-factor model. However, some further modifications are necessary or the resulting portfolio will typically be rather unbalanced, meaning one will be led to invest in a very small number of stocks each with a very large weight (in absolute value). To see this take the related objective of maximizing (5) subject to (3) only and where cs_i is given by (11). With very high probability all cs_i will be distinct so the resulting portfolio will then go long one unit in the stock with the highest cs_i and short one unit in the stock with smallest cs_i . Clearly, most hedge fund managers would shy away from such an

extremely unbalanced portfolio. Enforcing the additional constraint (4) will provide a certain amount of diversification but still not enough to arrive at portfolios with desirable return-risk properties in general.

Fortunately, there exist two widely-accepted approaches to achieve more balanced portfolios. For them to be well-defined, one needs to fix the sum of positive weights to a given value, say one. To do so, denote

$$w_i^+ = w_i \cdot \mathbb{1}\{w_i > 0\} \quad \text{and} \quad w_i^- = -w_i \cdot \mathbb{1}\{w_i < 0\} . \quad (12)$$

We then impose $\sum_{i=1}^T w_i^+ = 1$ which, under (3), implies that $\sum_{i=1}^T w_i^- = 1$ as well.

The first approach consists of using an upper bound for the (absolute) portfolio weights:

$$|w_i| \leq c, \quad \text{for all } i, \quad \text{for some } c > 0 . \quad (13)$$

For example, choosing $c = 0.1$ ensures that no stock in the long portfolio can have a weight greater than 10%, and similarly no stock in the short portfolio can have a weight greater than 10% either. In particular, the overall long-short portfolio invests in at least twenty stocks.

The second approach consists of introducing a penalty term for the estimated portfolio variance and thereby accounts for the extent to which the satisfaction scores of stocks in the portfolio co-vary.¹⁰ Denote by $\hat{\Sigma}$ a suitable estimator of the $N \times N$ covariance matrix of the returns of the N stocks in the investment universe and let $\mathbf{w} = (w_1, \dots, w_N)'$. Then the estimated portfolio variance is equal to

$$\hat{\sigma}^2(\mathbf{w}) = \mathbf{w}' \hat{\Sigma} \mathbf{w} . \quad (14)$$

Instead of maximizing (5) with respect to \mathbf{w} , one now maximizes a ‘penalized’ measure of portfolio customer satisfaction with respect to \mathbf{w} , which is given by

$$\sum_{i=1}^N w_i \cdot cs_i - \lambda \cdot \mathbf{w}' \hat{\Sigma} \mathbf{w} , \quad (15)$$

where $\lambda \in [0, \infty)$ expresses the severeness of the penalty for a large (estimated) portfolio variance.

Often times, fund managers even combine these two approaches in real-life portfolio formations. The portfolio optimization problem, in its most general form, is then expressed as

¹⁰Ideally, one would like to introduce a penalty for the true portfolio variance, but this is not feasible.

follows:

$$\begin{aligned}
\text{Maximize: } & \sum_{i=1}^N w_i \cdot cs_i - \lambda \cdot \mathbf{w}' \widehat{\Sigma} \mathbf{w} & (16) \\
\text{subject to: } & \sum_{i=1}^N w_i = 0, \quad \sum_{i=1}^N w_i^+ = 1, \quad |w_i| \leq c \text{ for all } i \\
& \sum_{i=1}^N w_i \cdot \text{beta}_i = \sum_{i=1}^N w_i \cdot \text{log-size}_i = \sum_{i=1}^N w_i \cdot \text{btm}_i = \sum_{i=1}^N w_i \cdot \text{mom}_i = 0
\end{aligned}$$

This is recognized as a quadratic programming problem and standard optimization software can be employed to find the optimal weight vector, denoted by \mathbf{w}^* . By choosing $c = 1$, the upper bound on the absolute portfolio weights can be dropped. Similarly, by choosing $\lambda = 0$, the penalty on the (estimated) portfolio variance can be dropped.

3.4 Data and Empirical Results

ACSI scores are downloaded from <http://www.theacsi.org>. We use the same list of firms as [Jacobson and Mizik \(2009b\)](#).¹¹ Corresponding data on stock returns, market size, and book-to-market are downloaded from CRSP.¹² Stock returns and market size are downloaded at monthly frequency; book-to-market is downloaded at yearly frequency. Book value is released once per year as part of fiscal year-end accounting disclosures, therefore updating book-to-market more frequently than annually would not make sense, and is not generally done in the finance literature.

Portfolios are formed at the beginning of a specific month always and will be updated on either a quarterly or yearly basis (details to follow shortly). Next, we need to be specific about how the inputs beta_i and mom_i for the portfolio optimization problem (16) are obtained for a specific firm i . Let us assume that the day of portfolio optimization is (say) 01/01/2001. Then beta_i is obtained by a time series regression of the stock returns of firm i on the returns on the S&P 500 index using the previous 60 months of data, i.e., the returns from 01/1996 until 12/2000. Furthermore, mom_i is obtained as the geometric average of the stock returns of firm i during the previous 12 months, excluding the most recent month; in our example it is the geometric average of the 11 returns from 01/2000 until 11/2000 (the practice of leaving the last month out to compute momentum is well established in the finance literature; for example, see [Asness \(1997\)](#)). Information from the last month should be left out so as to avoid the short-term (one-month) mean-reversion effect documented by [Lehmann \(1990\)](#).

¹¹We are grateful to Natalie Mizik for sharing with us the corresponding list of PERMNO firm identifiers.

¹²Source: CRSP[®], Center for Research in Security Prices. Graduate School of Business, The University of Chicago. Used with permission. All rights reserved. <http://www.crsp.com>.

To finalize the optimization process, we also need to describe how the estimated covariance matrix $\widehat{\Sigma}$ in the objective function of the portfolio formulation (16) is obtained. Following established practice, we use the shrinkage estimator of [Ledoit and Wolf \(2003\)](#) based on the previous 60 months of stock return data for all the N firms in the investment universe.

In the empirical application we consider both yearly updating in January and quarterly updating after the releases of the most recent ACSI scores, that is, in March, June, September and December. Until June 2010, ACSI scores were released according to this quarterly schedule and since then the releases occur according to a new monthly schedule. Only this current (monthly) schedule can be found on the ACSI website. The mapping back to the old quarterly schedule was communicated to us by [VanAmburg \(2011\)](#), director of ACSI, as follows:

- February = data now released in December and February
- May = data now released in April, May, and June
- August = data now released in July, August, and September
- November = data now released in October and November

Our portfolio formation period is 01/1997 until 12/2009 for yearly updating and 12/1996 until 11/2009 for quarterly updating, respectively, for a total of $T = 156$ out-of-sample monthly returns in each case.

At any given point in time, the investment universe consists of the firms in the database for which all relevant information is available. In particular, this requires a current ACSI score and also the ACSI score for the previous calendar year (so that the change in the ACSI score can be computed). It also requires a complete 60-month history of previous stock return data to compute beta and momentum and of course we also need information on market size and book-to-market. Between 1997 and 2009, the size of the investment universe ranges from $N = 50$ to $N = 74$, with the average size being 63.5 and the median size being 70.

Previous studies have shown that there are no significant short-term price movements around the release day of new ACSI scores; see [Fornell et al. \(2006\)](#) and [Ittner et al. \(2009\)](#). As a consequence, if there is any benefit from loading on stocks whose firms enjoy high customer satisfaction, it must come from mid-term or long-term portfolios. In order to examine both horizons, we use both quarterly and yearly updating of our long-short portfolios. A further motivation to also consider yearly updating is the concern of transaction costs.¹³ If a portfolio based on yearly updating delivers similar performance compared to a portfolio based on quarterly updating before transaction costs, then it will deliver better performance once transactions are factored in. [Ittner et al. \(2009\)](#), for example, do not find any meaningful

¹³This concern is even more pronounced for long-short portfolios compared to long-only portfolios, since the former incur (roughly) twice the transaction costs.

differences in portfolios that are updated on a yearly basis compared to portfolios that are updated on a quarterly basis (before transaction costs).

As is clear from the discussion in the preceding subsections, there are an infinite number of possible portfolio formation rules. First, there are several choices for the customer satisfaction scores cs_i . Second, there is continuum of choices for the input parameters c and λ in the portfolio optimization formulation (16). To keep the number of rules considered to a reasonable minimum, yet at the same time appropriately span the parameter space, we consider the following eight choices. For the customer satisfaction scores cs_i we consider the definition (6) in the spirit of Aksoy et al. (2008) and Jacobson and Mizik (2009b) and also the definition (11) with $\rho \in \{0, 0.5, 1\}$. Recall that $\rho = 0$ means that customer satisfaction is measured only in percentage changes (as in Ittner et al., 2009) and $\rho = 1$ means that customer satisfaction is measured only in levels. For the upper bound for the maximum weight for any stock in the long-short portfolio, we consider $c = 0.1$ and for the penalty for the portfolio variance we consider $\lambda \in \{0, 2 \cdot 10^3\}$. The value of $c = 0.1$ is reasonable as few (if any) fund managers would be willing to invest more than 10% of a portfolio (be it long or short) in any single stock. The second range of values, that is, for λ , was chosen by trial-and-error to yield portfolios that are significantly more diversified compared to the choice $\lambda = 0$. Another way of achieving more diversification would be to choose a smaller value of c , such as $c = 0.05$. However, this is not feasible for our particular application, as in certain years no solution for the formulation (16) can then be found at all.¹⁴

For any given portfolio, we report the following summary measures computed from the resulting $T = 156$ out-of-sample returns. First, the sample mean together with a corresponding t -statistic. Second, the (annualized) sample Sharpe ratio together with a corresponding t -statistic. The (annualized) sample Sharpe ratio, based on the observed returns r_1, \dots, r_T , is defined as

$$\widehat{\text{SR}} = \sqrt{12} \cdot \frac{\bar{r}}{s_r} \quad \text{with} \quad \bar{r} = \frac{1}{T} \sum_{t=1}^T r_t \quad \text{and} \quad s_r^2 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2. \quad (17)$$

We do not subtract the risk-free rate in the numerator of $\widehat{\text{SR}}$, since our portfolios are long-short rather than long-only. If anything, this introduces a bias in favor of finding mispricing of customer satisfaction, since it will lead to slightly larger values of the ratio. Moreover, the Sharpe ratio, compared to the raw sample mean of the returns, is arguably of greater concern to a fund manager. It also lends itself to somewhat easier interpretation as any number above 0.5 starts to become ‘interesting’ to a fund manager. As a reference, the CRSP value-weighted

¹⁴For $c = 0.05$, an investment universe of size at least 40 would be needed without the four neutrality constraints with respect to beta, size, book-to-market, and momentum. But with these constraints in place, a larger universe will generally be needed; in certain years our the size of the investment universe is as low as 50.

index (including distributions) on the S&P 500 universe of stocks has a Sharpe ratio of 0.5 over the 60-year period 1950–2010. Investment strategies whose Sharpe ratios are below 0.5 are usually not deemed economically significant enough to attract meaningful amounts of capital.

Crucially, we also provide two t -statistics for hypotheses on regarding expected returns and Sharpe ratios:

$$H_0 : \mathbb{E}(r_t) = 0 \quad \text{and} \quad H_0 : \text{SR} = 0, \quad \text{with } \text{SR} = \frac{\mathbb{E}(r_t)}{\text{SD}(r_t)}. \quad (18)$$

In terms of hypothesis testing, it does not really matter whether we base the test on the mean return or on the Sharpe ratio. This is because the true Sharpe ratio SR is positive if and only if the true mean $\mathbb{E}(r_r)$ is positive.¹⁵ Therefore, the two respective t -statistics should actually be very close to each other. To studentize we use HAC standard errors based on the prewhitened QS kernel with the automatic choice of bandwidth of [Andrews and Monahan \(1992\)](#).

In the tables that follow we also report the median size of the long portfolio (denoted by MSL) and the median size of the short portfolio (denoted by MSS) over the investment period. In this context, ‘size’ refers to the number of stocks (firms) in a portfolio.

3.4.1 Implementation Details

To be completely transparent about how we form our portfolios and compute the corresponding out-of-sample returns, a few remarks are in order.

It is, in principle, possible that outliers in the input data to the optimization formulation (16) lead to somewhat distorted portfolio weights. We circumvent this problem by properly ‘truncating’ very small and very large observations in any cross-sectional data set. This is often called ‘Winsorization’, a method that is widely used by quantitative portfolio managers; for example, see [Chincarini and Kim \(2006, p. 180\)](#).

Consider a set of numbers a_1, \dots, a_N . We first compute a robust measure of location that is not (heavily) affected by potential outliers. To this end we use the trimmed mean of the data with trimming fraction $\eta \in (0, 0.5)$ on the left and on the right. This number is simply the mean of the middle $(1 - 2\eta) \cdot 100\%$ of the data. More specifically, denote by

$$a_{(1)} \leq a_{(2)} \leq \dots \leq a_{(N)} \quad (19)$$

the ordered data (from smallest to largest) and denote by

$$M = \lfloor \eta \cdot N \rfloor \quad (20)$$

the smallest integer less than or equal to $\eta \cdot N$. Then the trimmed mean with trimming

¹⁵This would, of course, not be true if we subtracted the risk-free rate in the numerator of SR.

fraction η is defined as

$$\bar{a}_\eta = \frac{1}{N - 2M} \sum_{i=M+1}^{N-M} a^{(i)} . \quad (21)$$

We employ the value of $\eta = 0.1$ in practice.

We next compute a robust measure of spread. To this end we use the mean absolute deviation (MAD) given by

$$\text{MAD}(a) = \frac{1}{N} \sum_{i=1}^N |a_i - \text{med}(a)| , \quad (22)$$

where $\text{med}(a)$ is the sample median of a_1, \dots, a_N .

We finally compute upper and lower bounds defined by

$$a_{lo} = \bar{a}_{0.1} - 5 \cdot \text{MAD}(a) \quad \text{and} \quad a_{up} = \bar{a}_{0.1} + 5 \cdot \text{MAD}(a) . \quad (23)$$

The motivation here is that for a normally distributed sample, it will hold that $\bar{a} \approx \bar{a}_{0.1}$ and $s(a) \approx 1.5 \cdot \text{MAD}(a)$, where \bar{a} and $s(a)$ denote the sample mean and the sample median of a_1, \dots, a_N , respectively. As a result, for a ‘well-behaved’ sample, there will usually be no points below a_{lo} or above a_{up} . Our truncation rule is then that any data point a_i below a_{lo} will be changed to a_{lo} and any data point a_i above a_{up} will be changed to a_{up} . We apply this truncation rule to the data sets $\{\text{beta}_i\}$, $\{\log\text{-size}_i\}$, $\{\text{btm}_i\}$, and $\{\text{mom}_i\}$. We also apply it to the past stock return data (one observation period at a time) used to compute $\hat{\Sigma}$. (Of course, we do not apply this truncation rule to *future* stock return data used to compute portfolio out-of-sample returns.)

While outliers can be dealt with by the trimming procedures just described, there is potentially also the problem of missing future stock returns. Say, in the context of yearly updating, we form a portfolio on 01/01/2001 to be held throughout the calendar year of 2001. It is, in principle, possible that a firm who is included in the portfolio will be delisted during 2001. This delisting can either be due to good news associated price appreciation (such as a takeover) or by bad news associated with a price drop (such as bankruptcy). As a result, there will be some missing stock returns for such a firm. When this occurs we simply use the risk-free rate as a suitable replacement for the corresponding months of missing stock return data.¹⁶ In the absence of further information, it is reasonable to assume that the last price reflects most (if not all) of the economic impact of the news that caused the delisting, which justifies using the T-bill rate going forward, as is common practice in the finance literature.

¹⁶As measured by the 3-month T-bill rate; corresponding data was downloaded from CRSP.

3.4.2 Results for Yearly Updating

The results for yearly updating are presented in Table (1). None of the t -statistics are above 1.0 and some are even negative, though very small in magnitude. In each case, as it should be, the t -statistic for the sample average closely matches the t -statistic for the sample Sharpe ratio. The overall conclusion is, therefore, that no evidence for mispricing of customer satisfaction can be found.

One might argue that statistical significance is not necessarily the same as economic significance; however, no economic significance turns up either, as the largest observed (annualized) Sharpe ratio is below 0.25. As an aside, the largest t -statistics as well as the largest observed Sharpe ratio are found for the measure of customer satisfaction cs_i as in (11) with $\rho = 1$. Recall that $\rho = 1$ means that customer satisfaction is based only on the levels. This finding seems to contradict the argument of Ittner et al. (2009) for only using changes but not levels; see Subsection 2.1.

These results will perhaps come as a disappointment to the proponents of mispricing of customer satisfaction.¹⁷ A possible challenge on their part might be that yearly updating is too infrequent and that it results in buying some quite-past ‘winners’ whose stock price moved up before they were included in the long portfolio; and similarly in selling some quite-past ‘losers’ whose stock price moved down before they were included in the short portfolio. To examine the validity of such a challenge, we can examine the performance of alternative portfolios that use *future* customer satisfaction data. Such a strategy is of course not feasible in practice yet it does allow us to construct a ‘best case scenario’ for the value of customer satisfaction information (as measured by the ASCI at least). It is interesting to see the extent to which one would benefit if one had perfect foresight with respect to the next wave of ACSI scores (during they calendar year in which one will hold the portfolio). Such a strategy then corresponds to: include future ‘winners’ in the long portfolio and future ‘losers’ in the short portfolio.

The corresponding results from the ‘perfect foresight’ strategy are presented in Table (2). Sample means, sample Sharpe ratios, and t -statistics generally all increase compared to the feasible strategy of using past ACSI data. Nevertheless, not a single significant t -statistic can be found. Moreover, the largest (annualized) Sharpe ratio is below 0.32 such that no economic significance can be claimed either. If even ‘insider’ trading on future ACSI scores does not result in profitable portfolios, it seems difficult to make a convincing case for the mispricing of customer satisfaction.

Remark 3.1. Apparently, a related real-life trading strategy implemented by persons with access to ACSI scores two weeks before publication actually resulted in overall losses; for example, see <http://www.marketwatch.com/story/dont-let-scandal-hurt-michigan-survey>.

¹⁷Allow us to point out here that we have no personal stake at all in this debate.

3.4.3 Results for Quarterly Updating

The results for quarterly updating are presented in Table (3). As opposed to [Ittner et al. \(2009\)](#), we do get noticeable improvements compared to yearly updating.¹⁸ Nevertheless, we cannot (quite) find either statistical or economic significance. The largest t -statistic is around 1.8 and the largest (annualized) Sharpe ratio is around 0.47.

As previously discussed, one can argue that a t -statistic of 1.8 is significant at the 5% level if a one-sided test is carried out. Our response to this is two-fold. On the one hand, as mentioned earlier, we consider a two-sided test is more appropriate as economic theory is silent as to prior expectations on the returns. On the other hand, one should also not forget that we considered a total of eight investment strategies and that the 1.8 value is the largest of the eight corresponding t -statistics.¹⁹ If *any* adjustment is made at all for the implicit multiple testing scenario, then the 1.8 could certainly not be considered significant at the 5% level even if one-sided hypotheses are deemed appropriate.²⁰

In terms of economic significance, none of the strategies reach the threshold of 0.5 for the Sharpe ratio that we set by reference to the performance of the S&P 500; see Subsection 3.4. One could argue that the top-performing strategy, with a Sharpe ratio of 0.47, is ‘close’ to achieving economic significance, but once again we must point out that it is the best of eight strategies, and therefore it is artificially boosted by data-snooping biases; for example, see [Lo and MacKinley \(1990\)](#). Finally, and as with yearly updating, the largest t -statistics as well as the largest observed Sharpe ratio are found for the measure of customer satisfaction cs_i as in (11) with $\rho = 1$, based only on the levels. As noted earlier, this finding seems to contradict again the argument of [Ittner et al. \(2009\)](#) for only using changes but not levels; see Subsection 2.1.

4 Conclusions

Customers and customer relationships are considered key assets of the firm. Moreover, as noted in the Introduction, both academics and practitioners have come to view satisfied customers as a *sine qua non* for long term profitability. While it is relatively uncontroversial to suggest that satisfied customers buy more, engage in higher levels of repeat purchasing, and sometimes

¹⁸See Table 4 of [Ittner et al. \(2009\)](#). In their terminology, ‘long-term strategy’ corresponds to yearly updating while ‘short-term strategy’ corresponds to quarterly updating.

¹⁹Strictly speaking, we compute 16 t -statistics. For each portfolio formation rule there are two: one for \bar{r} and one for \widehat{SR} . Since these two values are basically the same for each rule there are really only eight ‘distinct’ t -statistics altogether.

²⁰The reader interested in modern methods to adjust for a multiple testing scenario is referred to [Romano and Wolf \(2005\)](#) and [Romano et al. \(2008\)](#).

spread positive word-of-mouth, there is far less unanimity about whether or not movements in customer satisfaction are related to mispricing in financial markets. In particular, an interesting and provocative stream of literature has emerged to investigate possible mispricing of customer satisfaction.²¹ Specifically, there are now quite a few published studies that examine the profitability of portfolio formation rules that exploit the (potential) mispricing of customer satisfaction. A key drawback, however, is that all these studies require the use of a risk model to analyze observed portfolio returns. Since the composition of the portfolios generally changes over time, the assumption of constant risk factor exposures, on which such models are based, is hardly tenable. Moreover, the direction of the induced bias is unclear; hence, both affirmative findings of mispricing and opposing findings of no mispricing will always be subject to criticism.

Our main contribution is therefore to suggest and develop alternative portfolio formation rules that alleviate the need for a risk model altogether. We demonstrated that this can be achieved by requiring our portfolios to be neutral with respect to all risk factors considered *a priori*. Therefore, no risk model is needed to evaluate the observed portfolio returns *a posteriori*. Our substantive conclusion is unambiguous: After considering a wide range of specific portfolio formation rules, and also two different updating frequencies, we fail to find any convincing evidence for mispricing of customer satisfaction. That is, there is no evidence for mispricing based on either statistical or economic significance.

Looking ahead, we see at least two fruitful avenues for future research. First, the methods and approach we present are general and can be applied to other contexts where possible mispricing is suspected. One would simply perform the maximization with respect to the new candidate marketing input (for example, brand value scores, and so on). Second, the construct ‘customer satisfaction’ has itself been subject to critique—as a sometimes less than informative predictor of future customer behaviors. It may well be the case that alternative measures of customer satisfaction (other than those utilized by the ACSI) are connected to mispricing. One prominent critic of standard approaches to measuring customer satisfaction is Frederic Reichheld, the author of the now popular and widely-used Net Promoter Score; see [Reichheld \(2003\)](#). That measure has been shown, in some contexts, to outperform customer satisfaction as a predictor of future customer behavior; it may therefore be worth investigating in the context of mispricing. We plan to address these issues in future research.

²¹For example, see *Marketing Science* (2009), 28(5).

References

- Aksoy, L., Cooil, B., Groening, C., Keiningham, T. L., and Yalçin, A. (2008). The long-term stock market valuation of customer satisfaction. *Journal of Marketing*, 72(4):105–122.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.
- Andrews, D. W. K. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60:953–966.
- Asness, C. S. (1997). The interaction of value and momentum strategies. *Financial Analysts Journal*, 53(2):29–36.
- Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.
- Chincarini, L. B. and Kim, D. (2006). *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*. McGraw-Hill, New York.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fornell, C., Mithas, S., and III, F. V. M. (2009). The economic and statistical significance of stock returns on customer satisfaction. (Commentary). *Marketing Science*, 28(5):820–825.
- Fornell, C., Mithas, S., Morgeson III, F. V., and Krishnan, M. S. (2006). Customer satisfaction and stock prices: High returns, low risk. *Journal of Marketing*, 70(1):3–14.
- Ittner, C., Larcker, D., and Taylor, D. (2009). The stock market’s pricing of customer satisfaction. (Commentary). *Marketing Science*, 28(5):826–835.
- Jacobson, R. and Mizik, N. (2009a). Customer satisfaction-based mispricing: Issues and misconceptions. (Rejoinder). *Marketing Science*, 28(5):836–845.
- Jacobson, R. and Mizik, N. (2009b). The financial markets and customer satisfaction: Reexamining possible financial market mispricing of customer satisfaction. *Marketing Science*, 28(5):810–819.
- Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.
- Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics*, 105(1):1–28.

- Lewellen, J. and Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289–314.
- Lo, A. and MacKinley, C. (1990). Data snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3:431–468.
- Luo, X., Homburg, C., and Wieseke, J. (2010). Customer satisfaction, stock market recommendations, and firm value. *Journal of Marketing Research*, 47:1041–1058.
- McAlister, L., Srinivasan, R., and Kim, M. (2007). Advertising, research and development, and systematic risk of the firm. *Journal of Marketing*, 71(1):35–48.
- Mittal, V. and Frennea, C. (2010). Customer satisfaction: A strategic review and guidelines for managers. Working Paper 10-701, Marketing Science Institute.
- Mizik, N. (2011). Personal communication, e-mail. (May 26).
- Newey, W. K. and West, K. D. (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.
- O’Sullivan, D., Hutchinson, M. C., and O’Connell, V. (2009). Empirical evidence of the stock market’s (mis)pricing of customer satisfaction. *International Journal of Research in Marketing*, 26:154–161.
- Reichheld, F. (2003). The one number you need to grow. *Harvard Business Review*, 81(12):46–54.
- Romano, J. P., Shaikh, A. M., and Wolf, M. (2008). Formalized data snooping based on generalized error rates. *Econometric Theory*, 24(2):404–447.
- Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.
- Tuli, K. and Bharadwaj, S. (2009). Customer satisfaction and stock returns risk. *Journal of Marketing Research*, 73:184–197.
- VanAmburg, D. (2011). Personal communication, e-mail. (June 6).

λ	\bar{r}	$\widehat{\text{SR}}$	MSL	MSS
cs _i as in (6)				
0	-0.000 (-0.015)	-0.004 (-0.015)	13	12
$2 \cdot 10^3$	-0.000 (-0.116)	-0.033 (-0.115)	22	21
cs _i as in (11) with $\rho = 0$				
0	-0.000 (-0.019)	-0.006 (-0.019)	13	12
$2 \cdot 10^3$	-0.000 (-0.040)	-0.1345 (-0.040)	19	22
cs _i as in (11) with $\rho = 0.5$				
0	0.002 (0.678)	0.161 (0.662)	13	12
$2 \cdot 10^3$	-0.000 (-0.005)	-0.002 (-0.005)	21	20
cs _i as in (11) with $\rho = 1$				
0	0.003 (0.929)	0.231 (0.922)	13	12
$2 \cdot 10^3$	0.002 (0.758)	0.185 (0.762)	23	20

Table 1: Summary statistics for various long-short portfolios based on portfolio formulation (16) with $c = 1$, using yearly updating, from 01/1997 until 12/2009. \bar{r} denotes the sample average of the $T = 156$ out-of-sample returns. $\widehat{\text{SR}}$ denotes the corresponding (annualized) Sharpe ratio, without subtracting the risk-free rate. t -statistics for these two summary statistics appear in parentheses below. MSL denotes the median size of the long portfolio over the investment period. MSS denotes the median size of the short portfolio over the investment period. In this context, ‘size’ refers to the number of stocks (firms) in a portfolio.

λ	\bar{r}	$\widehat{\text{SR}}$	MSL	MSS
cs_i as in (6)				
0	0.003 (1.138)	0.311 (1.135)	12	13
$2 \cdot 10^3$	0.002 (0.894)	0.231 (0.875)	24	21
cs_i as in (11) with $\rho = 0$				
0	0.002 (0.647)	0.199 (0.660)	13	12
$2 \cdot 10^3$	0.001 (0.604)	0.176 (0.620)	20	21
cs_i as in (11) with $\rho = 0.5$				
0	0.002 (0.726)	0.209 (0.733)	13	13
$2 \cdot 10^3$	0.002 (0.994)	0.282 (1.027)	24	21
cs_i as in (11) with $\rho = 1$				
0	0.002 (0.592)	0.141 (0.589)	13	13
$2 \cdot 10^3$	0.002 (0.978)	0.230 (0.967)	24	20

Table 2: Summary statistics for various long-short portfolios based on portfolio formulation (16) with $c = 0.1$, using yearly updating, from 01/1997 until 12/2009. \bar{r} denotes the sample average of the $T = 156$ out-of-sample returns. $\widehat{\text{SR}}$ denotes the corresponding (annualized) Sharpe ratio, without subtracting the risk-free rate. t -statistics for these two summary statistics appear in parentheses below. MSL denotes the median size of the long portfolio over the investment period. MSS denotes the median size of the short portfolio over the investment period. In this context, ‘size’ refers to the number of stocks (firms) in a portfolio. In contrast to Table (1), future values of cs_i are used in the portfolio formation; therefore, such a strategy is actually not feasible in practice.

λ	\bar{r}	$\widehat{\text{SR}}$	MSL	MSS
cs _i as in (6)				
0	0.002 (0.604)	0.174 (0.607)	12	12
$2 \cdot 10^3$	0.002 (1.024)	0.270 (1.016)	23	23
cs _i as in (11) with $\rho = 0$				
0	0.000 (0.063)	0.016 (0.062)	12	13
$2 \cdot 10^3$	0.001 (0.249)	0.061 (0.238)	20	22
cs _i as in (11) with $\rho = 0.5$				
0	0.002 (0.642)	0.184 (0.638)	13	13
$2 \cdot 10^3$	0.003 (1.704)	0.428 (1.665)	24	22
cs _i as in (11) with $\rho = 1$				
0	0.003 (1.128)	0.304 (1.151)	13	13
$2 \cdot 10^3$	0.004 (1.804)	0.464 (1.779)	22	21

Table 3: Summary statistics for various long-short portfolios based on portfolio formulation (16) with $c = 1$, using quarterly updating, from Q4/1996 until Q4/2009. \bar{r} denotes the sample average of the $T = 156$ out-of-sample returns. $\widehat{\text{SR}}$ denotes the corresponding (annualized) Sharpe ratio, without subtracting the risk-free rate. t -statistics for these two summary statistics appear in parentheses below. MSL denotes the median size of the long portfolio over the investment period. MSS denotes the median size of the short portfolio over the investment period. In this context, ‘size’ refers to the number of stocks (firms) in a portfolio.