

A New Portfolio Formation Approach to Mispricing of Marketing Performance Indicators: an Application to Customer Satisfaction

David R. Bell · Olivier Ledoit · Michael Wolf

Published online: 1 October 2014
© Springer Science+Business Media New York 2014

Abstract The mispricing of marketing performance indicators (e.g., brand equity, churn, and customer satisfaction) is an important element of arguments in favor of the financial value of marketing investments. Evidence for mispricing can be assessed by examining whether or not portfolios composed of firms that load highly on marketing performance indicators deliver excess returns. Unfortunately, extant portfolio formation methods that require the use of a risk model are open to the criticism of time-varying risk factor loadings due to the changing composition of the portfolio over time. This is a serious critique as the direction of the induced bias is unknown. As an alternative, we propose a new method and 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. We apply our method to a frequently studied marketing performance indicator, customer satisfaction, and 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.

Keywords Customer satisfaction · Financial performance · Long-short portfolio · Mispricing

1 Introduction

Executives and marketing academics alike believe that marketing investments contribute positively to the financial health and value of the firm.¹ Firms with strong brands, low rates of churn, and satisfied customers might therefore be expected to outperform their competitors not only in terms of sales and market share, but also in terms of long-term financial value as well. One form of compelling evidence for this relationship is the existence of mispricing, i.e., that marketing performance indicators provide additional information about the future health of the firm that is not immediately compounded into its stock price.

Several recent articles test for the relationships between financial performance (in the form of abnormal returns) and marketing investments in advertising quality, research and development efforts, and customer satisfaction (see, e.g., [12, 17, 18, 26]). The basic approach taken by these and related papers that examine mispricing is to (1) decide on a specific portfolio formation rule, (2) use this rule on the past data in order to observe corresponding returns, and (3) examine whether the observed returns are ‘abnormally’ good, as measured by statistical and economic 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 that load highly on the desired marketing performance indicator. Second, one can form long-short portfolios by buying stocks of firms with high

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

D. R. Bell (✉)
Marketing Department, The Wharton School, Philadelphia,
PA 19104-6340, USA
e-mail: davidb@wharton.upenn.edu

O. Ledoit · M. Wolf
Department of Economics, University of Zurich, CH-8032 Zurich,
Switzerland

O. Ledoit
e-mail: olivier.ledoit@econ.uzh.ch

M. Wolf
e-mail: michael.wolf@econ.uzh.ch

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

loadings while, at the same time, selling stocks of firms with low loadings. (As a practical matter, one also needs to determine the ‘right’ way to measure and characterize marketing performance and use this metric to construct the portfolios to be tested.) After deciding on a particular portfolio formation rule and observing the resulting returns, one is then left to judge whether those returns are indeed ‘abnormally’ good. In practice, 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.

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 (e.g., [5, 7]) 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. A serious problem with this approach is that it assumes the exposures to the various risk factors are constant over time. Unfortunately, this assumption is not fulfilled for portfolios formed on typical market performance indicators which are time-varying, since by definition, this causes the portfolio composition to change over time.

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 this is not the case as even long-short portfolios will likely 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.

Our contribution is twofold. First, we develop and implement a novel approach that eliminates the need to use a risk model altogether. This is important because the direction of the bias induced by the use of risk models on time-varying portfolios cannot be determined *ex ante*. In constructing our approach, we adopt the perspective of a real-world fund manager aiming to form a long-short portfolio based on marketing performance indicators, but 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*. Our contribution is quite general and likely of interest to other researchers who measure the financial performance consequences of various

marketing actions, including investments in brand equity, product quality, advertising effectiveness, customer satisfaction, and so on.

Second, we apply our new method to the most widely-studied source of possible mispricing, customer satisfaction. While academics have examined the financial performance consequences of reputation, brand image, and product quality, work on customer satisfaction is not only the most widespread but also 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 in terms of mispricing. A review article by [19] (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, findings on the latter point—whether or not there is mispricing of customer satisfaction—are decidedly mixed. Starting with [9], there have been several articles on whether the (potential) mispricing of customer satisfaction can be exploited to form investment portfolios with abnormal returns. This point is nicely made by [19] (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.

After implementing our new portfolio method which does not suffer from biases inherent in the application of risk models to mispricing, and analyzing a wide range of scenarios, our substantive contribution is the following: 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 firm-level satisfaction scores and their temporal evolution (at least as measured by the popular ACSI) are fully accounted for by the market.

The remainder of the paper is organized as follows. Section 2 provides more details on portfolio formation for tests of mispricing and reviews some previous findings on possible mispricing of customer satisfaction in particular. Section 3 provides the rationale for our new approach to portfolio formation, as well as the mathematical details. We also report the findings from the application of our method to the possible mispricing of customer satisfaction. Finally,

Section 4 concludes the paper with a brief discussion of the two contributions and implications for future research.

2 Background on Portfolio Evaluation and Application to Customer Satisfaction

In this section, we begin by reviewing the classic approach to portfolio formation and assessment of returns. In addition, we consider the efficacy of the approach in identifying abnormal returns to marketing performance, through the lens of recent debate on the mispricing of customer satisfaction. It is natural that many articles in the marketing literature that assess financial returns to marketing in general, and examine mispricing in particular, rely on the four-factor model in [5] as shown below in Eq. (1) and discussed shortly. The interpretation of model parameters is well established and the model embodies a straightforward test for the presence of abnormal returns. Unfortunately, a clean implementation of the approach remains elusive, as evidenced by a healthy debate on the mispricing of customer satisfaction.

The specific strand of literature on mispricing that motivates our own application of our new method dates back to [9] 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.² While evidence in favor of mispricing of this sort adds heft to the notion that marketing effort has important financial consequences, there are two main reasons why these findings have not been considered definitive by the field. First, the portfolio formation criteria have been critiqued as arbitrary and have been suspected as potential in-hindsight maximizers (see, e.g., [10]). Second, some of the findings are largely descriptive; ideally, outperformance should be backed up with a demonstration of statistical significance. Consequently, more recent studies tend to use more clear-cut portfolio formation rules and also employ statistical tests of significance in assessing portfolio outperformance.

The challenge of constructing portfolios and associated testing methods that offer unambiguous answers to the question of whether marketing drives (abnormal) financial returns is best seen through examples and reference to prior work. Next, we delineate long-only and long-short portfolio construction, and again, for ease of exposition illustrate our points with reference to the findings on mispricing of customer satisfaction.

²A large literature considers a number of financial consequences resulting from customer satisfaction, including financial risk, e.g., [26], analyst recommendations, e.g., [17], and related issues. We focus exclusively on mispricing and therefore limit our review to relevant articles in this stream of work.

2.1 Long-Only Portfolios

Recently, [12] attempted a replication of a previous study of [1]. Specifically, they constructed long-only portfolios based on both the absolute level of customer satisfaction of firms in the portfolio, as well as recent changes in customer satisfaction. Researchers who are interested in the mispricing of customer satisfaction can obtain satisfaction scores (on a 0–100 scale) via free download 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 a quarterly schedule; since then a monthly schedule has been in use.

In their article, [12] form long-only portfolios as follows. They 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. This yields the following four portfolios:

	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

While there is no direct economic theory per se, if customer satisfaction really is a leading indicator of financial performance then one might expect that Portfolio 1 will perform the best while Portfolio 4 will perform the worst. Furthermore, Portfolios 2 and 3 constitute an intermediate 'gray area'; see [1].

Now, since all four portfolios are long-only, and therefore, expected by construction to produce a positive return, mispricing cannot be tested by simply checking whether the average return is significantly different from zero. Instead, one needs to focus on the intercept in a suitable risk model. Unsurprisingly, [12] employ the standard and widely adopted [5] four-factor risk model:

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

Here, $\text{Ret}_{p,t}$ denotes the return of portfolio p during month t ; $\text{Ret}_{r_f,t}$ 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’.

The difficulties with this standard approach are best understood through the detailed findings in [1] and [12]. Both studies 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, which leads to changes in the composition of the portfolios as some firms are improving and others declining with respect to their performance on the ACSI scores.³ Nevertheless, the main focus of these and similar studies is the sign and significance of $\hat{\alpha}_1$.

Aksoy et al. [1] obtained a t -statistic of 2.49 and [12] find a smaller value of only 1.82; both studies report point estimates of around 0.005. Jacobson and Mizik [12] speculated 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 respective standard errors of $\hat{\alpha}_1$ were computed in the two studies. Aksoy et al. [1] used the heteroskedasticity and autocorrelation (HAC) standard errors of [21], without specifying their choice of bandwidth, while [12] do not say.⁴

A more minor issue with respect to the substantive finding, i.e., whether or not customer satisfaction information is subject to mispricing, is the appropriate number of tails for the test. The t -statistic of [12] 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 [12] used a two-sided test while [8] argued 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. We prefer a two-sided test as the a priori view on the sign of α_1 is based on previous empirical findings (using largely overlapping time periods)

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

⁴As an aside, we are somewhat surprised at the popularity that the sub-optimal [21] 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 [2] and [3].

rather than on economic theory. None of the other $\hat{\alpha}_p$ estimates are established as significant, though $\hat{\alpha}_4$ does have a negative sign in both studies.

Further substantive evidence on whether mispricing is present is given by [22] who reexamine the specific trading strategy of [9] 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. [9] found very attractive properties of the top-quintile portfolio, but only provide descriptive measures. O’Sullivan et al. [22], on the other hand, failed to find statistical significance. They consider two investment periods: February 1997 until May 2003, which corresponds to the investment period of [9], and March 1996 until May 2006, which corresponds to the investment period of [12]. The two resulting t -statistics for $\hat{\alpha}_p$ in Eq. (1) for this ‘top quintile’ portfolio are 0.73 and 0.84, respectively.⁵ Finally, [22] also examine a trading strategy closely related to the one of [12]. The t -statistic for $\hat{\alpha}_p$, for the sample period February 1996 until May 2006, is 0.93.

The troubling lack of agreement among all these studies and a potential driver is clearly identified by [12]. There, the authors note 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. Of course it is natural that firm-level customer satisfaction scores contain sufficient temporal variation to alter the portfolios from one period to the next. Again, we are focused on the mispricing of satisfaction information for ease of exposition only; it is clear that portfolio formation based on any other marketing metric that varies over time would induce the same kind of problem. Candidate metrics of interest in the field include brand equity, customer churn rates, and social media activity.

Hence, in these contexts, it is unrealistic to assume that the coefficients α_p , β_p , γ_p , and κ_p remain constant over time. The failure of this key assumption thus renders the all the above findings on mispricing of satisfaction questionable to some extent. Furthermore, it is important to note that, in principle, the failure of the assumption can bias t -statistics upwards or downwards, depending on circumstances. Thus, there is no straightforward consequence of this model-based problem for the answer to the substantive question of interest: Namely, are marketing actions capable of delivering abnormal financial returns?

⁵As alternatives to the four-factor risk model, [22] also employ the market model and the three-factor risk model of [7]. The resulting t -statistics for $\hat{\alpha}_p$ are then even smaller in these cases.

As a potential remedy to this general problem [12] also consider an alternative approach that allows for time-varying risk factors β_p , γ_p , and κ_p , using a methodology proposed by [15]. After employing this procedure to mispricing of satisfaction, 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 [12], specifically, to allow for time-varying risk factors based on methodology proposed by [18].

We, therefore, conclude that studying the potential mispricing of marketing performance indicators (customer satisfaction in the current example), 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. Returning again to the mispricing of satisfaction in particular, we note that [1] constructed 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

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 [12] 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 [20].

⁷They call the resulting portfolio ‘High – Low’ instead.

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, 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 the marketing performance indicator.

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 [5]:

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

Here, $\text{Ret}_{LS,t}$ denotes the return of the long-short portfolio in period t and the right hand side regressors are defined as in Eq. (1). Using this formulation, [1] found a point estimate of $\hat{\alpha}_{LS} = 0.0088$ for the mispricing of satisfaction, 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 Eq. (2).

While this does not get to the heart of the econometric issues, there is nevertheless potential merit in using more sophisticated measures of the marketing performance indicators themselves. The ability of a model to detect mispricing is necessarily linked to the way in which the marketing variables influence the construction of portfolios. An example of this in the literature on mispricing of satisfaction is the approach to long-short portfolio construction in [10]. 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 (i.e., up vs. down movement), they use the actual change expressed as a percentage, arguing 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, and might therefore lead to a more efficient delineation of portfolios, which would aid in detection of mispricing.⁸

Following this approach, [10] 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

⁸Nevertheless, it may be difficult to make a compelling case for a specific form of measurement for the marketing performance indicator in question. Jacobson and Mizik [11], for example, stated that ‘No one right way and single formation criteria exist to form portfolios that can be assessed for mispricing’.

every quarter after new ACSI scores are released. Ittner et al. [10] used daily return data and consider holding periods of 365 days, 180 days, and one quarter (i.e., until the next wave of ACSI scores is released). In total, they compute six *t*-statistics for their long-short portfolios: There are three holding periods and two parameters of interest (the raw expected return and the intercept of risk model).⁹ All six *t*-statistics lie between 0.72 and 0.83. Obviously, these values are in stark contrast to the findings of [1]. While [10] used a bigger universe of firms (243) compared to [1], who use 151 firms and a slightly longer investment period¹⁰, these differences hardly seem capable of explaining the large discrepancies in *t*-statistics.

One is, therefore, left to speculate whether the way [10] use the ACSI scores in portfolio formation (i.e., using only the percentage change) is somehow less informative compared to the way [1] use them (i.e., using both levels and the signs of the changes). In short, we have identified two key issues that hinder estimation of mispricing for marketing performance indicators: (1) the time-varying composition of the portfolios formed and the concern that the estimate of mispricing is therefore time-varying as well, and (2) potential deficiencies in the measurement approach taken when the marketing performance indicators are used to construct the portfolios themselves.

3 A new Portfolio Formation Approach

The essence of our new approach is conceptually and methodologically straightforward and also informed by how a fund manager might trade on marketing performance indicators that were leading metrics for financial performance. Specifically, 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 [5]. In other words, we aim to be market-neutral (i.e., β_{LS} equal to zero), size-neutral (i.e., γ_{LS} equal to zero), book-to-market neutral (i.e., δ_{LS} equal to zero), and momentum-neutral (i.e., 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 Eq. (2) if all four risk factors loadings are equal to zero. In

other words, by properly accounting for the four risk factors a priori (i.e., 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 vexing issue of time-varying coefficients. Note too 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 (e.g., 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 simultaneously satisfied. To see how this works, 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

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

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

In addition, 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

$$\begin{aligned} \sum_{i=1}^N w_i \cdot \text{beta}_i &= 0, \quad \sum_{i=1}^N w_i \cdot \text{log-size}_i = 0, \\ \sum_{i=1}^N w_i \cdot \text{btm}_i &= 0, \quad \text{and} \quad \sum_{i=1}^N w_i \cdot \text{mom}_i = 0. \end{aligned} \tag{4}$$

3.2 Loading on Marketing Performance Indicators

Any portfolio satisfying Eqs. (3)–(4) would be a valid choice for our purposes in the sense of being a portfolio that

⁹Ittner et al. [10] employed the three-factor model of [7].

¹⁰They 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.

is long-short and neutral with respect to the four risk factors. But so far no information on the marketing performance indicator has been built in. Let mp_i denote a specific measure of marketing performance, e.g., customer satisfaction, brand equity, customer retention, and so on, for company i . We then aim, over all ‘valid’ portfolios, to maximize the performance measure of the portfolio expressed as

$$\sum_{i=1}^N w_i \cdot mp_i . \tag{5}$$

Needless to say, given the examples for customer satisfaction, the form of the marketing performance measure also requires careful consideration. One possibility, in the spirit of [1] and [12], in the customer satisfaction literature is to take the sum of two indicator functions: whether the marketing performance score of company i is above the national average and whether the last change has been positive:

$$mp_i = \mathbb{1}\{\text{score}_i > \text{national average}\} + \mathbb{1}\{\text{score}_i > \text{previous score}_i\} . \tag{6}$$

If we apply this approach to the customer satisfaction example we have employed throughout this paper, i.e., $mp_i = cs_i$, then we get $cs_i = 0$ for firms in Portfolio 4, $cs_i = 1$ for firms in Portfolios 2 and 3, and $cs_i = 2$ for firms in Portfolio 1. On the other hand, the proposal of [10] corresponds to

$$mp_i = \frac{\text{score}_i}{\text{previous score}_i} - 1 . \tag{7}$$

Given that prior research demonstrates that it is not only the time-varying composition of portfolios that makes identification of mispricing a challenge, but also the measurement approach for the marketing performance indicator as well, we introduce a more general measure of marketing performance that includes Eq. (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 not be appropriate. The two scales can be made comparable by the standard approach of converting the two quantities to z -scores (i.e., 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

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

Then, the z -score corresponding to a_i is defined as

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

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

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

where the subscript l stands for ‘level’ and the subscript c stands for (percentage) ‘change’. A more general measure of the marketing performance indicator under consideration, e.g., customer satisfaction, is then

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

This includes Eq. (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 marketing performance measure Eq. (5) subject to the constraints Eqs. (3)–(4). 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 Eq. (5) subject to Eq. (3) only and where mp_i is given by Eq. (11). With very high probability, all mp_i will be distinct so the resulting portfolio will then go long one unit in the stock with the highest mp_i and short one unit in the stock with smallest mp_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

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

We then impose $\sum_{i=1}^T w_i^+ = 1$ which, under Eq. (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:

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

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 20 stocks.

The second approach consists of introducing a penalty term for the estimated portfolio variance and thereby accounts for the extent to which the marketing performance scores of stocks in the portfolio co-vary.¹¹ 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 Eq. (5) with respect to \mathbf{w} , one now maximizes a ‘penalized’ measure of portfolio marketing performance 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 follows:

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

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 An Application to Customer Satisfaction

In order to demonstrate how our approach can be implemented, we return again to customer satisfaction. This has two advantages. First and foremost, it allows us to benchmark our method in the domain where mispricing of marketing performance has been most widely studied. Second, since our method: (1) does not suffer from the important

critiques relating to time-varying portfolio composition, and (2) uses a very general approach to marketing performance measurement, we are also able to contribute to the substantive debate about whether customer satisfaction information is mispriced.

We proceed by downloading the ACSI scores from <http://www.theacsi.org> and using the same list of firms as in [12].¹² Corresponding data on stock returns, market size, and book-to-market are downloaded from CRSP.¹³ 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) January 1, 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 January 1996 until December 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 January 2000 until November 2000 (the practice of leaving the last month out to compute momentum is well established in the finance literature; e.g., see [4]). Information from the last month should be left out so as to avoid the short-term (1 month) mean-reversion effect documented by [14].

To finalize the optimization process, we also need to describe how the estimated covariance matrix $\hat{\Sigma}$ in the objective function of the portfolio formulation (16) is obtained. Following established practice, we use the shrinkage estimator of [13] 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

¹¹Ideally, one would like to introduce a penalty for the true portfolio variance, but this is not feasible.

¹²We are grateful to Natalie Mizik for sharing with us the corresponding list of PERMNO firm identifiers.

¹³Source: 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>.

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 [27], 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 January 1997 until December 2009 for yearly updating and December 1996 until November 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 [9] and [10]. 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.¹⁴ 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. [10], for example, do not find any meaningful 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 [1] and [12], 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 [10], 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.¹⁵

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

$$\widehat{\text{SR}} = \sqrt{12} \cdot \frac{\bar{r}}{s_r} \quad \text{with} \quad \bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$$

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

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 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.

¹⁴This concern is even more pronounced for long-short portfolios compared to long-only portfolios, since the former incur (roughly) twice the transaction costs.

¹⁵For $c = 0.05$, 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.

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 regarding expected returns and Sharpe ratios:

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

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 $\mathbb{E}(r_t)$ is positive.¹⁶ Therefore, the two respective *t*-statistics should actually be very close to each other. To standardized, we use HAC standard errors based on the prewhitened QS kernel with the automatic choice of bandwidth of [3].

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 Eq. (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 [6, page 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 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.

¹⁶This would, of course, not be true if we subtracted the risk-free rate in the numerator of SR.

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

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

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 {beta_{*i*}}, {log-size_{*i*}}, {btm_{*i*}}, and {mom_{*i*}}. We also apply it to the past stock return data (one observation period at a time) used to compute $\bar{\Sigma}$. (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 January 1, 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.¹⁷ 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.

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.

¹⁷As measured by the 3-month T-bill rate; corresponding data was downloaded from CRSP.

Table 1 Summary statistics for various long-short portfolios based on portfolio formulation (16) with $c = 1$, using yearly updating, from January 1997 until December 2009

	λ	\bar{r}	\widehat{SR}	MSL	MSS
\bar{r} denotes the sample average of the $T = 156$ out-of-sample returns. \widehat{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	cs _i as in Eq. (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 Eq. (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 Eq. (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 Eq. (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	

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 Eq. (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 [10] for only using changes but not levels; see Section 2.1.

These results will perhaps come as a disappointment to the proponents of mispricing of customer satisfaction.¹⁸ 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 ASCI 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 ASCI 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 ASCI 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 ASCI scores

¹⁸Allow us to point out here that we have no personal stake at all in this debate.

Table 2 Summary statistics for various long-short portfolios based on portfolio formulation (16) with $c = 0.1$, using yearly updating, from January 1997 until December 2009

	λ	\bar{r}	\widehat{SR}	MSL	MSS
\bar{r} denotes the sample average of the $T = 156$ out-of-sample returns. \widehat{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	cs _i as in Eq. (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 Eq. (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 Eq. (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 Eq. (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	

two weeks before publication actually resulted in overall losses; for example, see <http://www.marketwatch.com/story/dont-let-scandal-hurt-michigan-survey>.

3.4.3 Results for Quarterly Updating

The results for quarterly updating are presented in Table 3. As opposed to [10], we do get noticeable improvements compared to yearly updating.¹⁹ 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 twofold. 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.²⁰ 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.²¹

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 Section 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 [16]. 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 Eq. (11) with $\rho = 1$, based only on the levels. As noted earlier, this finding seems to contradict again the argument of [10] for only using changes but not levels; see Section 2.1.

²⁰Strictly 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.

²¹The reader interested in modern methods to adjust for a multiple testing scenario is referred to [24] and [25].

¹⁹See Table 4 of [10]. In their terminology, ‘long-term strategy’ corresponds to yearly updating while ‘short-term strategy’ corresponds to quarterly updating.

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

	λ	\bar{r}	\widehat{SR}	MSL	MSS
\bar{r} denotes the sample average of the $T = 156$ out-of-sample returns. \widehat{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	cs _{<i>i</i>} as in Eq. (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>i</i>} as in Eq. (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>i</i>} as in Eq. (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>i</i>} as in Eq. (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	

4 Conclusions

The mispricing of marketing performance indicators is of theoretical and empirical interest to academics and practitioners alike. Of theoretical interest because evidence of mispricing would lend strong support to the idea that marketing performance measures are potentially ‘leading indicators’ of the financial health of a firm. This speaks to the *sine qua non* of marketing as a discipline that has a key role to play in long-term profitability. Many firms that invest in say brand and customer equity anticipate financial returns that are not only positive but also potentially superior, as a consequence. Unequivocal empirical support for mispricing would validate this belief, and also be of practical value to fund managers and investors as well.

It is, therefore, no surprise that a literature has emerged to examine the profitability of portfolio formation rules that exploit the (potential) mispricing of marketing performance indicators. 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 not tenable. More critically, 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 contribution is twofold. First, our main contribution is 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. Second, we weigh in on the mispricing debate for the most widely studied marketing performance indicator, customer satisfaction.²² Here, 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.

4.1 Future Research

Looking ahead, we see at least two fruitful avenues for future research. First, the methods and approach we present

²²For example, see *Marketing Science* (2009), 28(5).

are general and can be applied to other contexts where possible mispricing is suspected. As noted earlier, one would simply perform the maximization with respect to the new candidate marketing input (e.g., brand equity 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 [23]. 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.

References

- Aksoy L, Cooil B, Groening C, Keiningham TL, Yalçin A (2008) The long-term stock market valuation of customer satisfaction. *J Mark* 72(4):105–122
- Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59:817–858
- Andrews DWK, Monahan JC (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60:953–966
- Asness CS (1997) The interaction of value and momentum strategies. *Financ Anal J* 53(2):29–36
- Carhart M (1997) On persistence in mutual fund performance. *J Financ* 52(1):57–82
- Chincarini LB, Kim D (2006) Quantitative equity portfolio management: an active approach to portfolio construction and management. McGraw-Hill, New York
- Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J Financ Econ* 33(1):3–56
- Fornell C, Mithas S, III FVM (2009) The economic and statistical significance of stock returns on customer satisfaction. (Commentary). *Mark Sci* 28(5):820–825
- Fornell C, Mithas S, Morgeson FVIII, Krishnan MS (2006) Customer satisfaction and stock prices: High returns, low risk. *J Mark* 70(1):3–14
- Ittner C, Larcker D, Taylor D (2009) The stock market’s pricing of customer satisfaction. (Commentary). *Mark Sci* 28(5):826–835
- Jacobson R, Mizik N (2009a) Customer satisfaction-based mispricing: Issues and misconceptions. (Rejoinder). *Mark Sci* 28(5):836–845
- Jacobson R, Mizik N (2009b) The financial markets and customer satisfaction: Reexamining possible financial market mispricing of customer satisfaction. *Mark Sci* 28(5):810–819
- Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J Empir Financ* 10(5):603–621
- Lehmann BN (1990) Fads, martingales, and market efficiency. *Q J Econ* 105(1):1–28
- Lewellen J, Nagel S (2006) The conditional CAPM does not explain asset-pricing anomalies. *J Financ Econ* 82(2):289–314
- Lo A, MacKinley C (1990) Data snooping biases in tests of financial asset pricing models. *Rev Financ Stud* 3:431–468
- Luo X, Homburg C, Wieseke J (2010) Customer satisfaction, stock market recommendations, and firm value. *J Mark Res* 47:1041–1058
- McAlister L, Srinivasan R, Kim M (2007) Advertising, research and development, and systematic risk of the firm. *J Mark* 71(1):35–48
- Mittal V, 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 WK, West KD (1987) A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708
- O’Sullivan D, Hutchinson MC, O’Connell V (2009) Empirical evidence of the stock market’s (mis)pricing of customer satisfaction. *Int J Res Mark* 26:154–161
- Reichheld F (2003) The one number you need to grow. *Harv Bus Rev* 81(12):46–54
- Romano JP, Shaikh AM, Wolf M (2008) Formalized data snooping based on generalized error rates. *Economet Theor* 24(2):404–447
- Romano JP, Wolf M (2005) Stepwise multiple testing as formalized data snooping. *Econometrica* 73(4):1237–1282
- Tuli K, Bharadwaj S (2009) Customer satisfaction and stock returns risk. *J Mark Res* 73:184–197
- VanAmburg D (2011) Personal communication, e-mail. (June 6)