MARKETING SCIENCE
Vol. 28, No. 5, September–October 2009, pp. 820–825
ISSN 0732-2399 | EISSN 1526-548X | 09 | 2805 | 0820

DOI 10.1287/mksc.1090.0505

Commentary

The Economic and Statistical Significance of Stock Returns on Customer Satisfaction

Claes Fornell
Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109,
cfornell@umich.edu

Sunil Mithas
Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742,
smithas@rhsmith.umd.edu

Forrest V. Morgeson III
National Quality Research Center, Ross School of Business, University of Michigan,
Ann Arbor, Michigan 48109, morgeson@umich.edu

According to Jacobson and Mizik [Jacobson, R., N. Mizik. 2009. The financial markets and customer satisfaction: Reexamining possible financial market mispricing of customer satisfaction. *Marketing Sci.* **28**(5) 810–819], excess stock portfolio returns for firms with strong customer satisfaction are small and statistically insignificant, and if there is any above-market performance at all, it is due to a small set of firms in the computer and Internet industries. But their data seem to suggest the opposite. The returns are actually both exceptionally large and significant. Using monthly data, their portfolio consisting of strong American Customer Satisfaction Index (ACSI) firms outperformed the market by 0.0053, corresponding to 6.4% cumulative risk-adjusted above-market returns on an annual basis over a 10-year period—a performance that would beat at least 99% of all large-cap U.S. stock funds tracked by Morningstar. Using a different treatment of risk, their annualized risk-adjusted return is a whopping 8.4% better than market. After eliminating computer, Internet, and utility companies, they find that the monthly risk-adjusted abnormal returns drop to 0.0045, which corresponds to an annual above-market return of 5.4%. This too is better than 99% of all actively managed stock funds in the population. Yet Jacobson and Mizik conclude that these returns are not statistically significant and that there is no evidence that stock returns from firms with strong customer satisfaction outperform the market over the long run.

The failure to reject the null hypothesis is probably due to a lack of statistical power in Jacobson and Mizik's analysis. We discuss why this is likely the case and then present new data updating the results from our original article [Fornell, C., S. Mithas, F. Morgeson III, M. S. Krishnan. 2006. Customer satisfaction and stock prices: High returns, low risk. *J. Marketing* **70**(1) 3–14]. The above-market returns persist and are both economically and statistically significant.

Key words: customer satisfaction; stock prices; stock returns; risk; market value; stock portfolios
History: Received: March 10, 2009; accepted: March 11, 2009; processed by Jean-Pierre Dubé.

Introduction

By our count, Jacobson and Mizik (2009; J&M hereafter) perform 64 statistical tests of alpha—some theoretically motivated, others not. Three turn out to be "significant"—which is about what one would expect by chance. But failing to find a statistically significant alpha that would lie in the extreme tail of the population of alphas from all actively managed funds seems strange. Without indicating that their returns are actually very large, J&M conclude that there are no statistically significant risk-adjusted above-market returns from their portfolio of relevance (portfolio 1: firms with strong and improving customer satisfaction). Yet above-market, risk-adjusted annualized returns over a 10-year period amount to 6.4% (Table 2 of J&M, panel A) or 8.4% (Table 2 of J&M, panel B), depending on the treatment of risk. Either way, the probability of finding another large-cap U.S. stock fund with equal or better performance is next to zero. Of the 1,709 funds tracked by Morningstar, 99% do not even come close (http://www.Morningstar.com). In contrast, J&M's results for the portfolio with weak and declining customer satisfaction (portfolio 4 in Table 2 of J&M, panel A) display negative abnormal returns of −1.4% on an annual basis over 10 years. The difference in portfolio returns based on strong versus weak customer satisfaction is thus very sizeable.

In some contexts, substantive significance may differ from statistical significance if the sample size is small, the time period short, or if the data exhibit highly unusual variation. None of this is true here, however. The time period is long. The returns on customer satisfaction are not volatile. On the contrary, they tend to have lower volatility, lower systematic risk (Fornell et al. 2009, 2006), and lower downside risk (Tuli and Bharadwaj 2009). If excess returns persist over time—say, 10 years—it is, by definition, difficult to make the case that they are due to chance or to compensation for risk. Pure luck rarely lasts that long, and if the risk factors did not have an adverse effect, their impact, if any, must have been small. Failure to reject the null hypothesis under these circumstances is more likely due to low *p*-values relative to the power of the test.

The lack of power in J&M's analysis has to do with the statistical model used as well as certain steps J&M take that have a detrimental effect on statistical power. A case in point is the application of two-tailed tests (as J&M use in all their tests) when a one-tailed test is called for. The theory and the substantive hypothesis posit that strong customer satisfaction is associated with above-market returns and that weak customer satisfaction is associated with the opposite. The hypothesis is *not* that firms with strong customer satisfaction have negative abnormal returns. Accordingly, when a one-tail, rather than a two-tail, test is applied to J&M's results, the above-market risk-adjusted returns are indeed significant at the 5% level. But even so, one should be careful to not overinterpret statistical significance or the lack thereof. Even if alpha was not statistically significant, it does not follow that there is no evidence for above-market returns. Similar to O'Sullivan et al. (2009), the data presented by J&M suggest that the probability of above-market returns because of customer satisfaction is much higher than the probability that the returns are equal to or worse than market. If the odds are even 10 to 1 that a portfolio would provide excess returns over a 10-year period versus a market return, there is little doubt what a rational investor would do.

Statistical power is further diluted by selective trimming of the data. Subsequent to the statistical testing of the full portfolio, J&M eliminate those firms with the strongest relationship between customer satisfaction and stock returns (computer and Internet firms) and also another group (utilities) with a more modest, but still sizeable positive relationship (about 2% risk-adjusted above-market annual returns). As a result, the sample size is smaller, which of course has a negative impact on statistical power (because of higher idiosyncratic volatility). But there is a more serious problem with selective (nonrandom) trimming of data. Under certain circumstances it might be possible to use inferential theory (while not without caveats) for analyzing the eliminated samples separately, but under no circumstances would it be possible to draw on statistical probability theory with respect to the original sample once it has been tampered with. Obviously, taking out observations from the right tail of the distribution weakens relationships and makes it "easier" not to reject the null hypothesis, but eliminating observations from the left tail would have the opposite effect and "restore" significance. Proceeding in this manner, one can obtain whatever results one wants, but it has little to do with statistical hypothesis testing.

Hence, J&M's reported returns are actually both large *and* statistically significant for the complete and uncorrupted sample. However, let us nevertheless consider—for the purposes of examining *substantive* significance—J&M's trimmed sample (without utility, computer, and Internet firms). According to J&M's results (in their Table 2), the annual risk-adjusted above-market cumulative returns drop about 1% (to 5.4% or 5.1%, depending on which model is used). Even though this is quite close to the full sample return and would still place the portfolio in the top 1% among all large-cap U.S. stock portfolios over a 10-year period (with due consideration given to differences in how risk is (not) compensated for in the Morningstar data and different market returns over a small portion of the nonoverlapping time period), it is not statistically significant according to J&M.

As for J&M's portfolio of computer and Internet firms, its returns are off the charts: annualized risk-adjusted above-market returns of 32.4% (Table 2 of J&M, panel A) or 38.4% (Table 2 of J&M, panel B) over a 10-year period. Let us put this in perspective. A 2%–3% annual above-market return over a 10-year period is highly unusual (even without risk adjustment) and considered strong (Banjo 2008, Barras et al. 2009). But according to J&M, the 32.4% above-market returns are only significant at the 5% level—not the 1% level—and should presumably be viewed with some caution. As the probability of finding a stock fund with better performance is zero, this type of reasoning seems odd. What we have here is not really a diversified portfolio suitable for capital asset pricing models (CAPM) (there are 10 computer and Internet firms divided into four portfolios). It is not particularly meaningful to use a broad market proxy (and its associated risk factors) as a reference portfolio, as it appears that J&M have done. The lack of diversification would produce very large standard errors of the estimates and further dilute the power of the test.

There is yet one more power problem. It is more vexing because it is inherent in the CAPM as well as the extensions by Fama and French (1996) and Carhart (1997). The test statistics (that J&M rely on) have very low power in detecting abnormal returns. As demonstrated by the findings of Kothari and Warner (2001), the misspecification is so severe that just about any stock fund's performance assessment is unreliable. For example, above-market annual risk-adjusted returns even as high as 3% are not detected almost 70% of the time, and the problem gets even worse when the number of securities in a portfolio is less than 75. J&M do not report the number of firms in each of their portfolios, but it is certainly much less than 75.

J&M do not report fit statistics either, but it is important that the variance of the regression residual is small. Fama and French (1996) report R^2s above 0.9. We suspect that the residual variance in J&M's models is much larger with lower R^2s. In part, this is because the idiosyncratic volatility is not diversified away. Because the standard error of the intercept (alpha in this case) is directly proportional to the standard error of the residuals of the underlying four-factor regression models that J&M use, this is a serious concern. But even with sufficient diversification, we have never been able to fit any version of the CAPM particularly well to customer satisfaction returns. It is not that these models do not fit from the standpoint of traditional statistical criteria (i.e., a significant F-statistic), but they have a residual too large to be consistent with the efficient markets theory assertion that the risk factors explain all variation in returns.

In the remaining analyses (reported in their Table 3), J&M cite Lewellen and Nagel (2006) as justification for estimating time-varying risk factor models, but Lewellen and Nagel actually show that betas do not vary enough to explain asset-pricing anomalies and that time-varying risk factor models perform "nearly as poorly as the unconditional CAPM" (p. 289). Lewellen and Nagel also demonstrate that if the conditional (time varying) model holds, there should only be small deviations from the standard model (much smaller than what has been observed empirically and probably much smaller than found by J&M). Even though it seems that the betas have thus been inflated, it still turns out that the risk-adjusted abnormal returns of J&M's portfolio of relevance (portfolio 1) are high, so we see no reason to comment further, except to point out that in our original article (Fornell et al. 2006) we too use time-variant betas, but in a different context and with a different purpose. We will present more on that here, but before we get to that discussion, let us briefly summarize our earlier findings and also bring in new and updated empirical evidence.

In our 2006 study, we reported that customer satisfaction, as measured by the American Customer Satisfaction Index (ACSI), is significantly related to market value of equity, yet 5- and 15-day event studies showed that news about ACSI results does not move share prices. This apparent inconsistency was the catalyst for examining whether a customer satisfaction-based trading strategy might provide excess stock returns. We presented results from two stock portfolios. Both were constructed with ACSI data: a hypothetical back-tested portfolio and an actual portfolio with real investments. Both portfolios produced large above-market returns. The back-tested portfolio generated a cumulative return of 40%—about three times better than the S&P 500 at 13%—over a period of about seven years. However, back testing has obvious shortcomings. Some consider all such work meaningless because it capitalizes on chance due to data mining (Black 1993). Professional investors, in particular, seem to discount the value of back-tested strategies unless they are also combined with real returns—ex post, it may not be all that difficult to come up with a winning strategy.

We therefore also presented results from an actual stock portfolio. Its cumulative return was +75%, compared with −19% for the S&P 500. At the risk of stating the obvious, a few points should be kept in mind. These returns are sizeable. They are also large relative to market. They were not obtained from developing markets or from investments in distressed companies, speculation in commodity prices, or similarly risky investments, but from big-brand, generally large-cap consumer companies with substantial market shares and strong customer relationships. The returns were not a result of investments in small stocks or high book-to-market stocks. Beta risk was low. Idiosyncratic risk was low. Downside risk was low. Since the publication of our study, other researchers have confirmed many of its conclusions. Aksoy et al. (2008) and O'Sullivan et al. (2009) report large above-market returns for strong customer satisfaction portfolios (although the latter use an approach similar to J&M and do not find statistical significance). Ngobo et al. (2009) find that customer satisfaction is associated with higher earnings per share and better analyst forecasts. Tuli and Bharadwaj (2009) show that positive changes in customer satisfaction lead to lower volatility in stock returns and lower systematic risk. Anderson and Mansi (2009) report that customer satisfaction has value-relevant information incremental to accounting measures by studying the relationship between ACSI, credit ratings, and cost of debt financing. Similarly, Chen et al. (2008) find that ACSI provides information incremental to return on assets.

Table 1 Four-Factor Asset Pricing Model

	Daily returns	Monthly returns
α	0.054***	1.014***
t-value	3.179	3.559
Annualized α (%)	13.6	12.2
Adjusted R^2	0.575	0.590

Notes. Daily returns on ACSI portfolio, 05/01/2000 through 01/30/2009 ($n = 2{,}201$). Monthly returns on ACSI portfolio, 05/2000 through 01/2009 ($n = 105$).
***$p < 0.01$.

New Data

Between 2000 and 2004, the actual, real-money long-short portfolio described in Fornell et al. (2006) consistently generated above-market returns. Did this performance continue in subsequent years? If alpha was not statistically significant, one would predict not. Let us therefore update the portfolio through January 2009 and use the same four-factor CAPM used by J&M. Even though our past experience with CAPM and its extensions has not indicated a good fit to customer satisfaction data, for purposes of comparison, it may be instructive to find out what this model implies with respect to these returns. Table 1 shows abbreviated results based on the value-weighted composite market return from Kenneth French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html).

Despite the low power of the model, the returns are all statistically significant; but contrary to efficient markets theory, they are not completely accounted for by the risk factors. More than 40% of the variance in the returns is unexplained. The estimates of alpha are fairly close for monthly and daily data with annualized above-market risk-adjusted returns of 12.2% and 13.6%, respectively. As before, these returns are also considerably better than any of the large-cap funds tracked by Morningstar over roughly the same period of time, but this is not an entirely fair comparison because few mutual funds do short selling. A better comparison might be found among hedge funds, but relevant data are more difficult to come by. Most hedge funds also do not last for more than five years, and although there may well be hedge funds that have done better, we are not aware of any in a comparable class of assets.

Capitalizing on Marketing Knowledge: Alpha Picking and Beta Surfing

Much empirical research in marketing seeks data from consumer buyers because buyers have information that the company does not have about attitudes, intentions, awareness, knowledge, preferences, and customer satisfaction. With the help of this information, the firm presumably becomes better equipped to compete in product markets. What our findings suggest is that buyers in product markets may hold information of relevance for financial markets as well. That is, buyers, in the aggregate, may have information not yet impounded in share prices.

Let financial economics debate the relevance and substantiation of efficient markets theory. As long as there is no agreement on the definition of risk, excess returns are assumed to be constant, and a general absence of theory in the CAPM extensions (Carhart 1997, Fama and French 1996), there will be no end to this debate. Marketing science has little to contribute and little to gain from participating in it. On the other hand, it would not be a stretch to suggest that marketing—the bridge between buyers and sellers, between supply and demand, and between production and consumption—could contribute to value-relevant knowledge. In fact, it would be surprising if it could not. As buyers gain more power, primarily because of globalization and availability of information, at the expense of sellers, assets of supply (balance sheet assets) become less predictive about future wealth creation. Market-based assets, on the other hand, become more predictive (for a discussion, see Fornell 2007). A good deal of knowledge about these assets—how to measure, how to price, and how to grow them—resides in marketing. Instead of trying to accommodate financial capital asset pricing models with all their assumptions, low power, and problems of empirical fit, marketing science might well be capable of making its own contribution to alpha, beta, and to the concept of risk. The common assumption in finance that buyers of equity (investors) have homogenous views of risk would be foreign to marketing. What is seen as risk factors for Fama-French (size and book-to-market) and Carhart (momentum) may not at all be proxies for firm distress (see also Chung et al. 2006). Even beta as a measure of systematic risk might be challenged. A low beta is generally considered favorable because it suggests low exposure to systematic risk. Yet investors would obviously prefer a low (or negative) beta in down markets and a high positive beta in up markets.

The idea of separating alpha and beta has recently become popular among institutional investors (Bernstein 2007, *Economist* 2008). The purpose is to lock in the market return (beta) with low-cost index funds and seek alpha elsewhere. But alpha is very hard to come by in a $30 trillion worldwide stock/bond market, so many funds actually deliver market returns or less. Customer satisfaction appears to deliver excess returns via both alpha and beta—a quality that may contribute to the difficulty of capital asset pricing models in explaining the returns. Much of the performance of the ACSI portfolio in 2006, for example, was due to a beta that surged well above

Table 2 Portfolio Betas

Year	Beta of ACSI portfolio	S&P 500 annual return (%)
2000	0.727	−12.24
2001	0.723	−13.04
2002	0.811	−23.37
2003	0.878	26.38
2004	0.843	8.99
2005	0.958	3.00
2006	0.980	13.62
2007	0.700	3.53
2008	0.515	−38.49
	Low	High
Beta range		
Down markets	0.515	0.811
Up markets	0.700	0.980
Beta mean		
Down markets	0.694	
Up markets	0.872	

one during the latter half of the year. Good beta surfing is characterized by a higher coefficient in up-markets and a lower one in down-markets. This is illustrated in Table 2, albeit in a simplified summary, which provides annual betas for the portfolio along with S&P 500 performance. Betas in down-markets range from 0.515 to 0.811, compared to 0.700 to 0.980 in up-markets. The importance of this can hardly be overstated. For example, in the recessionary market collapse of 2008, the portfolio was only exposed to about 52% of the market loss.

The CAPM does not recognize a surfing beta as a deliberate strategy for seeking excess returns but rather as a risk to be considered. However, if a company's satisfied customers are the last to leave and the first to return, as customer satisfaction theory would suggest (reflected in earnings protection in down-markets and a surge in up-markets), beta itself (in addition to alpha) might well be a vehicle for excess returns.

Aside from the lack of empirical fit to demand-based assets, there are many problematic issues surrounding the original CAPM and its lack of explanatory power. The empirically based extensions by Fama and French (1996) and Carhart (1997) tend to have better empirical fit but lack theory. Although Fama and French (2004) do not consider brute empiricism fatal (p. 39), it seems doubtful—especially in view of the 2008–2009 stock market collapse—just how much these types of models have really contributed to our understanding of how markets work and how assets are priced. Perhaps it would be more constructive for marketing to look toward its own work and expertise. For example, the capital asset pricing models that estimate the financial values of a firm's customers, based on discounted cash flow analysis, are by now well established in marketing (Gupta et al. 2004, 2006; Gupta and Lehmann 2003, 2006; Rust et al. 2000; Villanueva and Hanssens 2007; Wiesel et al. 2008; Winer 2001). The sum total of the values implied by these models should approximate the value of the firm's business operations. Although there is much that remains to be done, and predictions have sometimes been off, these models might prove to be a suitable option for pricing intangible market-based assets. Not only do they provide a possible foundation for the above-market returns that we report, they also suggest that the value of marketing information might be severely underpriced in financial markets.

References

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

Anderson, E. W., S. A Mansi. 2009. Does customer satisfaction matter to investors? Findings from the bond market. *J. Marketing Res.* **46**(5). Forthcoming.

Banjo, S. 2008. In this bear market, be ready to buy. *Wall Street Journal* (July 6) 1.

Barras, L., O. Scaillet, R. Wermers. 2009. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *J. Finance*. Forthcoming.

Bernstein, P. L. 2007. *Capital Ideas Evolving*. John Wiley & Sons, Hoboken, NJ.

Black, F. 1993. Beta and return. *J. Portfolio Management* **20**(1) 8–18.

Carhart, M. 1997. On persistence in mutual fund performance. *J. Finance* **52**(1) 57–82.

Chen, C. X., E. M. Matsumura, J. Y. Shin. 2008. The effect of competition on the contracting use of customer satisfaction: Evidence from the American Customer Satisfaction Index (ACSI). AAA 2008 MAS Meeting Paper, http://ssrn.com/abstract=1004117.

Chung, Y., H. Johnson, M. Schill. 2006. Asset pricing when returns are nonnormal: Fama-French factors versus higher-order systematic co-moments. *J. Bus.* **79**(2) 923–940.

Economist, The. 2008. Money for old hope. (March 1).

Fama, E. F., K. R. French. 1996. Multifactor explanations of asset pricing anomalies. *J. Finance* **51**(1) 55–84.

Fama, E. F., K. R. French. 2004. The capital asset pricing model: Theory and evidence. *J. Econom. Perspectives* **18**(3) 25–46.

Fornell, C. 2007. *The Satisfied Customer: Winners and Losers in the Battle for Buyer Preference*. Palgrave Macmillan, New York.

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

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

Gupta, S., D. R. Lehmann. 2003. Customers as assets. *J. Interactive Marketing* **17**(1) 9–24.

Gupta, S., D. R. Lehmann. 2006. Customer lifetime value and firm valuation. *J. Relationship Marketing* **5**(2/3) 87–110.

Gupta, S., D. R. Lehmann, J. A. Stuart. 2004. Valuing customers. *J. Marketing Res.* **41**(1) 7–18.

Gupta, S., D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, S. Sriram. 2006. Modeling customer lifetime value. *J. Service Res.* **9**(2) 139–155.

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

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

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

Ngobo, P. V., J. F. Casta, O. J. Ramond. 2009. Customer satisfaction and financial analysts earnings forecast errors. Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1348272.

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

Rust, R. T., V. A. Zeithaml, K. N. Lemon. 2000. *Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy*. Free Press, New York.

Tuli, K. R., S. G. Bharadwaj. 2009. Customer satisfaction and stock returns risk. *J. Marketing*. Forthcoming.

Villanueva, J., D. M. Hanssens. 2007. *Customer Equity: Measurement, Management, and Research Opportunities*. Now Publishers, Hanover, MA.

Wiesel, T., B. Skiera, J. Villanueva. 2008. Customer equity: An integral part of financial reporting. *J. Marketing* **72**(2) 1–14.

Winer, S. R. 2001. A framework for customer relationship management. *California Management Rev.* **43**(4) 89–105.