Do Hedge Fund Managers Have Stock-Picking Skills?*

Wesley R. Gray†
Empirical Finance, LLC
wes@empiricalfinancellc.com

Andrew E. Kern†
Empirical Finance, LLC
andy@empiricalfinancellc.com

This draft: November 10, 2009
First draft: August 1, 2008

Note: This paper is an adaptation of a broader academic work, synthesized for a practitioner audience.

---

†Mr. Gray and Mr. Kern are both in their final years of their finance Ph.D. programs. They attend the University of Chicago Booth School of Business and the University of Missouri, respectively. Wes and Andy are also Managing Members of Empirical Finance, LLC, an investment management organization designed to turn academic insights into investment performance. This document has been provided to you solely for information purposes and does not constitute an offer or solicitation of an offer, or any advice or recommendation, to purchase any securities or other financial instruments, and may not be construed as such.

*We would like to thank Dave Carlson, Hui Chen, John Cochrane, Lauren Cohen, Cliff Gray, Eugene Fama, Ron Howren, Carl Luft, Stavros Panageas, Shastri Sandy, Gil Sadka, Amir Sufi, Pietro Veronesi, Rob Vishny, anonymous referees for the Midwest Finance Association, and seminar participants at the University of Chicago Booth School of Business.

EXECUTIVE SUMMARY

Many academics and proponents of passive money management believe the active management industry provides no value-added, because fund managers lack stock-picking skills. This constituency believes markets are efficient; funds that outperform are considered lucky, funds that underperform are considered unlucky. End of story.

Unlike previous research, which studies broad cross-sections of portfolio returns from mutual funds, or analyzes the suspect data from the various hedge fund data providers, we perform a clinical study of investment recommendations submitted to Valueinvestorsclub.com (VIC). VIC is a private-access club with a membership base that consists primarily of smaller fundamentals-based hedge fund managers and their associates. The environment of VIC provides an excellent opportunity to assess the stock-picking skills hypothesis.

We find overwhelming evidence that the hedge fund managers in our sample have stock-picking skills. The average one-, two-, and three-year raw returns of long recommendations submitted to the site over the January, 1 2000 to December 31, 2008 period are 17.11%, 45.02%, and 74.39%, respectively. After controlling for market risk, size risk, book-to-market exposure, and momentum risk, via a control portfolio, we find excess one-, two-, and three-year returns of 9.52%, 19.03%, and 23.60%, respectively. These results are all highly statistically significant and are well in excess of any management and performance fees these managers charge their investors.

While the full sample results are certainly impressive and provide convincing evidence that some hedge fund managers have stock-picking skills, the more intriguing evidence comes from the analysis of VIC recommendations *after* controlling for the quality of a recommendation. In order to control for quality, we use the VIC rating system as a proxy for an idea's future outperformance. For example, an idea with a rating of "2" from the VIC community should produce less outperformance than an idea with a rating of "8."

What we find is that ratings are highly related to future outperformance, which is another manifestation of hedge fund manager stock-picking skills. In fact, the average raw return to ideas rated in the top 20% of all ideas submitted have one-, two-, and three-year raw returns of 27.76%, 46.59%, and 86.86%, respectively. Meanwhile, ideas rated in the bottom 20% of all ideas submitted have one-, two-, and three-year raw returns of 8.56%, 32.62%, and 46.25%, respectively. After controlling for risk, the highest rated ideas provide alpha of 21.69%, 26.15%, and 42.58% on a one-, two-, and three-year horizon, whereas, the lowest rated ideas provide alpha of -.16%, 4.21%, and -9.02%, over the same time periods. The data certainly suggest that the members of VIC have stock-picking skills.

Our primary conclusion is that there are certainly professional money managers in the marketplace that can provide value above and beyond a passively managed mutual fund. It still may be the case that investing in an index fund is the optimal decision for investors who lack the time and/or skills to identify skilled fund managers. However, the rewards of finding "good" managers, like those involved with VIC, are certainly high enough to justify the increased fees they likely charge. This provides preliminary evidence that the active money management industry's primary product isn't snake oil.

***

Professional active investors understandably resist the implication that, as a group, they are destined to fail in their attempts to outperform the market and are merely a drag on society. This has led to harsh criticism of the efficient market hypothesis and sometimes a rebuke of academic finance generally. Charles Munger, the co-chairman of Berkshire Hathaway and lifelong partner of Warren Buffett, summed up a familiar sentiment in a 2003 lecture at the University of California at Santa Barbara:

"First, he [the academic] said Berkshire beat the market in common stock investing through one sigma of luck, because nobody could beat the market except by luck. This hard-form version of efficient market theory was taught in most schools of economics at the time. People were taught that nobody could beat the market. Next the professor went to two sigmas, and three sigmas, and four sigmas, and when he finally got to six sigmas of luck, people were laughing so hard he stopped doing it" (Munger (2003)).

And yet, the preponderance of the evidence from academic research suggests investors have little to no stock picking skills on average. Moreover, French (2008) estimates that the "unskilled" active management industry cost investors over $100 billion in 2007. It would be a catastrophic failure for microeconomic theory if the relatively competitive active management industry was able to convince investors they provided services worth $100 billion, when in fact they provided little to no value beyond an index fund. In fact, it would be time for capitalists to throw the towel in and bring the socialists into power—because a centralized social planner could efficiently invest everyone's capital in a value-weighted index fund with relative low cost.

If we assume markets are perfectly efficient and active managers lack skill, we must simultaneously—and paradoxically—assume investors invested in active funds are completely irrational. Both the efficient market and the completely irrational investor hypotheses are far-fetched…but, if this is true, why haven't the academics found any evidence for stock-picking skill?

One of the primary issues with academic research is the consistent focus on portfolio returns. However, Cohen, Polk, and Silli (2009) make the argument that analyzing portfolio returns is not a test of stock-picking skill, or "value-added," because portfolio returns may disguise a fund manager's stock-picking ability. Their paper argues that managers have incentives to hold diversified portfolios that consist of their "best ideas" and other positions to "round out" their portfolio. Some reasons a manager may include zero-alpha positions in their portfolio are to decrease volatility, price impact, illiquidity, and regulatory/litigation risk. Berk and Green (2004) formalize aspects of this argument and point out that the very nature of fund evaluation may cause managers to hold many stocks which they have little conviction, since the managers may be punished for exposing their investors to idiosyncratic risk.

Berk and Green also conclude that research analyzing fund manager portfolio returns and/or persistence in returns says little about the skill level of managers, but is really a test of the efficiency of the capital allocation markets. Their basic argument is that "alpha" decreases with assets under management, and high alpha managers receive more and more assets under management until, at the margin, investors are indifferent between investing in an index fund and an active fund. For example, if a manager can produce 10% alpha on $50 million AUM, he will receive more AUM until his alpha is driven to almost 0%. So while this manager has provided great returns for his early investors, and provides adequate for his current investors, a researcher looking at his latest portfolio returns will conclude he has no alpha and has no skills.

An alternative approach to testing the stock-picking hypothesis, which does not suffer from the issues in studying portfolio returns, is to analyze individual recommendations from superstar managers or stock analysts. These studies confirm the hypothesis of no stock-picking skill from previous research. Desai and Jain (1995) examine the performance of recommendations made by "superstar" money managers and find little evidence of superior stock-picking skill. Barber et al. (2001) confirm this result and find that excess returns to the recommendations of stock analysts are not reliably positive.

The study of individual stock recommendations is certainly a step in the right direction for testing the stock-picking hypothesis. However, there are potential issues with testing the stock-picking hypothesis in the aforementioned studies. In the Desai and Jain and Barber et al. studies it is unclear why superstar managers or analysts would share profitable trading opportunities with the general public, so their results suggesting no stock-picking skills are not surprising.

Another angle on the stock-picking skill hypothesis has been to study the "smart money," which usually translates into studying hedge fund return databases. However, these papers are plagued because of data problems. First, hedge fund return databases suffer from survivorship bias (funds that go out of business are difficult to track), and self-selected reporting (managers may only report their returns to the hedge fund database creators when they have good performance) (Fung and Hsieh (2000)). Second, hedge fund managers sometimes hold illiquid assets or engage in return smoothing, which causes their reported hedge fund returns to exhibit large autocorrelations (Asness, Krail, and Liew (2001); Getmansky, Lo, and Makarov (2004)). Third, hedge fund database returns may be unreliable because the same hedge funds sometimes report different returns to different hedge fund database creators (Bollen and Pool (2008)). Fourth, hedge fund managers often hold assets which have payoffs that are option-like and non-linear. This payoff profile makes it difficult for researchers to assess hedge fund performance when they analyze hedge fund manager returns using traditional linear factor models (Fung and Hsieh (2001)). Finally, Griffin and Xu (2009) address the aforementioned issues with hedge fund return database biases, by analyzing hedge fund performance via their required 13F equity filings. The issue with Griffin and Xu's analysis is that they can only examine long-equity positions and ignore intraquarter trading.

Our data, the full sample investment recommendations shared on the private website Valueinvestorsclub.com (VIC), although imperfect, does not suffer from many of the biases found in previous hedge fund research addressing the no stock-picking skill hypothesis. Moreover, the proprietary data allow me to study individual fund manager recommendations as opposed to fund manager portfolios, which is likely a better setting to identify manager stock-picking skills.

Using two buy-and-hold-abnormal-return (BHAR) approaches, we find abnormal returns for long positions that are economically large and statistically significant across various holding periods. Evidence for stock-picking skills in short positions is directionally correct, but statistically inconclusive.

To further test the stock-picking hypothesis, we analyze the relationship between the average ratings VIC members assign to recommendations and the recommendation's ex-post abnormal returns. In line with the hypothesis that some fund managers have stock-picking skills, we find evidence that the investors in our sample are able to decipher which stocks will perform the best. This result holds for both long and short recommendations.

The remainder of the paper is organized as follows. Section I discusses the data. Section

II provides the main results on the characterization of value investor decisions in our sample. Section III tests for stock-picking skill via abnormal return analysis. Section IV examines the relation between ex-ante VIC idea ratings and ex-post abnormal returns, and section V concludes.

I. Data

A. *Value Investors Club*

The data in this study are collected from a private internet community called Valueinvestorsclub.com (VIC), an "exclusive online investment club where top investors share their best ideas."[1] The site has been heralded in many business publications as a top-quality resource for those who can attain membership (e.g. *Financial Times*, *Barron's*, *Business Week*, and *Forbes*).[2] The site was founded by Joel Greenblatt and John Petry, both successful value investors and managers of the large hedge fund Gotham Capital. The site was created with \$400,000 of start-up capital with the goal of being a place for "the best-quality ideas on the Web" (Barker (2001)). The investment ideas submitted on the club's site are broad, but are best described as fundamentals-based. The VIC website mentions that it is open to any well thought-out investment recommendation, but that it has a particular focus on equity or bond-based plays (either long or short), traditional asset undervaluation plays (high B/M, low P/E, liquidations, etc.) and investment ideas based on the notion of value as articulated by Warren Buffett (firms selling at a discount to their intrinsic value irrespective of common valuation ratios).

Membership in the club is capped at 250 and admittance to the club is based on an initial write-up of an investment idea. If the quality of the research is satisfactory and the aspiring member deemed a credible contributor to the club, he is admitted. Once admitted, members are required to submit two ideas per year with a maximum of six ideas per year—the maximum exists to ensure only the member's best ideas are submitted. Members share comments and rate each other's ideas on a scale of 1 (bad) to 10 (good). In addition, there is a weekly prize of \$5,000 awarded to the best idea submitted (VIC management determines the winner; community ratings have no bearing on who wins the prize). Members are monitored to ensure they submit at

---

[1] http://www.valueinvestorsclub.com/Value2/Guests/Info.aspx
[2] http://www.valueinvestorsclub.com/Value2/Guests/Info.aspx

least two credible ideas per year, and members failing to meet the high standards of the club are dismissed.

An important aspect of VIC is that members' identities are not disclosed to the general public or to the other members of the club. The intent behind this policy is to keep individual VIC members from forming outside sharing syndicates with selected members, who could then take their valuable comments and ideas away from the broader VIC community. In addition, the anonymity requirement ensures the message board does not become a venue for hedge fund managers to "signal" to potential investors or market their services to the general public.

Unfortunately, because membership of VIC is strictly confidential, we are unable to reveal detailed statistics on the subject. However, the management of VIC agreed to disclose that the preponderance of VIC members are long-focused fundamentals-based hedge fund managers who have small to mid-size assets under management ($10 million to $250 million).

B. *Data Description*

We analyze all investment reports submitted to VIC since the club's founding on January 1, 2000 through December 31, 2008. These reports represent *all* reports submitted to VIC over the entire time period the club has existed; ideas that subsequently do poorly are not dropped from the website and therefore the database we create does not suffer from an ex-post selection bias (although we cannot rule out disingenuous ex-post changes to the historical content of the data as described in the case of I/B/E/S analyst data by Ljungqvist et al. (2009)). In total we examine 3273 investment submissions. Report length can range from a few hundred to a few thousand words (see appendix for an example write-up). Investment ideas are wide-ranging with respect to the asset traded, where the asset is traded, and the complexity of the strategy employed.

For each investment report analyzed, we record various data: date and time of submission, symbol, price (at time of recommendation), market(s) traded, security(s) traded, strategy recommended (long, short, or long/short), and the "reasons for investing."

All data collected are unambiguous except for the "reasons for investing." We compile a list of sixteen investment criteria that are frequently cited in VIC submissions. Criteria were judged to be sufficiently common if at least 10 investment submissions acknowledged the use of the category. The sixteen categories are as follows: *lack of sell-side analyst coverage*, *tangible* *asset undervaluation* (high book-to-market, hidden real estate assets, etc.), *insider buying/selling* (Seyhun (1988)), *intrinsic value undervaluation* (e.g. discounted cash flow analysis, low P/E, EBIT/TEV, P/Sales, industry undervaluation, and hidden growth opportunities), *complicated business or taxes creating investor confusion*, *"sum-of-parts" discount*, *liquidation potential*, *active share repurchase programs* (Ikenberry et al. (1995)), *recent restructuring or spinoff situation*, *misunderstood net operating loss tax assets*, *merger arbitrage* (Mitchell and Pulvino (2001)), *stub arbitrage* (Mitchell, Pulvino, and Stafford (2002)), *activist involvement* (Boyson and Mooradian (2007)), *merger arbitrage trading opportunity*, *turnaround and/or bankruptcy emergence*, and *pair trade arbitrage* (Froot and Dabora (1999)).

We read every investment idea and assign it the appropriate categories. For example, the VIC submission cited in the appendix received four category labels: *tangible asset undervaluation*, *insider buying*, *intrinsic value undervaluation*, and *net operating loss tax assets*. By assigning investment submissions discrete criteria, we capture the essence of the why VIC members make their recommendations.

We then match the firms associated with a VIC recommendation to accounting and stock return data from CRSP/COMPUSTAT. For the purposes of this study, we only analyze US exchange-traded long and short common stock recommendations. We do not analyze U.S. common equity investment recommendations that have payoffs one may consider non-linear or inappropriate to analyze with linear factor asset pricing models because they would bias the results (Fung and Hsieh (2001)). Specifically, we eliminate all recommendations classified as merger arbitrage, stub arbitrage, pair-trade, and liquidation, long/short recommendations, non-common-equity ideas (e.g. options or preferred stock). We also eliminate foreign-traded/ADR recommendations.

Of the 3273 observations in the original sample, 2832 refer to U.S. securities. Of these 2832 observations, 2698 are recommendations on U.S. common stock securities. After the restrictions described above, we are left with 2066 U.S.-equity long recommendations and 252 U.S.-equity short recommendations.

We must further constrain our sample to those firms with contemporaneous data available from CRSP/Compustat. The extent to which sample sizes are reduced based on data requirements on CRSP/Compustat depend on the abnormal return calculation method employed. The sample with the requisite data to perform the control-firm BHAR analysis has 1671 long recommendations and 198 short recommendations. The sample with the necessary data to perform the benchmark-portfolio BHAR analysis consists of 1610 long recommendations and 198 short recommendations.

Finally, it is important to note that the ideas under analysis are the most simple, straightforward common equity recommendations submitted to VIC, and are further limited by the data available on CRSP/Compustat. The exclusion of the many complicated arbitrage trades and special situation scenarios submitted to VIC, but not analyzed due to data and analysis constraints, may bias the evidence. These sophisticated trades require advanced knowledge and understanding of niche securities and/or access to expensive resources (i.e. lawyers, industry specialists, and tax experts). In a Grossman and Stiglitz (1980) equilibrium where arbitrageurs are compensated for their information discovery efforts, one may hypothesize that these investments would have better gross returns (before costs of information collection) than situations requiring less effort. If this price discovery reward story is to be believed, the data under analysis will likely be biased and favor the null hypothesis that VIC members have no skill.

II. The Characteristics of Fundamental Value Investor Decisions

Grossman and Stiglitz (1980) argue that market prices can never be perfectly efficient. If prices were always efficient, skilled investors who acquire private information (via more efficient collection and processing of available information) would never be rewarded. And if skilled investors had no incentive to engage in the price discovery process, efficient market prices would be not be an equilibrium condition, but a simple case of extremely good luck.

Grossman and Stiglitz provide a compelling case that skilled investors, who are rewarded for their private information, are critical to an efficient price discovery process. However, there is little empirical evidence on where skilled investors look to generate their private information. Are these investors fixated on book value because the market cannot properly calculate asset value? Do they incorporate information from open market share repurchases, insider buying patterns, post earnings announcement drift, accruals, or use other alpha producing strategies documented in the academic literature? In this section, we examine how the value investors in our sample make investment decisions.

Using the full sample of recommendations (n=3273), we find that our sample of investors focus on US based common stock investments (82% of total recommendations), but find value in other markets as well: 13% of the recommended securities are internationally traded and .4% are non-equity investments. We also find that long recommendations in common stock represent the bulk of ideas submitted (86%) (see table 1, panel B).

Table 2 shows the frequency with which VIC members base their investment on various criteria.[3] We find that investors in our sample are overwhelmingly concerned with assessing intrinsic value: discounted cash flow models, earnings multiples, GARP (growth at a reasonable price), and other similar valuation techniques are used most frequently (87% include this analysis in their recommendation). However, nearly 24% of the recommendations do incorporate the classic value technique of focusing on tangible asset undervaluation. Other popular tools used by VIC members are open market repurchases (12%), the presence of net operating loss assets (5%), restructuring and spin-off situations (5%), and insider trading activity (5%).

VIC members often cite more than one criterion in an investment analysis. Table 3, panel A summarizes the frequency with which various combinations of criteria are cited. Panel A shows that although value investors are highly focused on intrinsic value, many cite additional criteria. This indicates that the skilled investors in our sample are not one-dimensional. Some of the most common criteria combinations paired intrinsic undervaluation with signaling factors such as share repurchase programs, insider buying, and activist involvement.

While the investors in our sample use a wide range of tools in their investment decisions, their analysis tends to focus on a defined set of criteria.[4] Table 3, Panel B shows that value investors typically employ up to three different criteria when making investment decisions. Ninety-eight percent of the recommendations cite three or fewer investment criteria, whereas only 2% cite four or more. We speculate that asset specific issues (e.g. a liquidation trade, by its nature, exclusively focuses on tangible asset undervaluation), specialization in a specific approach to deriving private information, and resource limitations are the primary reasons why the professional value investors in our sample focus on very few criteria when making investment decisions.

[3] I analyze the full sample in this section, however, the characteristics of the sub-samples I use for the asset pricing tests are very similar.

[4] I conjecture that recommendations represent a majority of the investment thesis and that members do not hold back information.

In table 4 we present more detailed descriptive statistics of the securities recommended segregated by type of recommendation (long versus short). In panel A we tabulate the sector classification: The recommendations are weighted towards manufacturing firms, representing 32.38% (40.91%) of the total of long (short) recommendations. Other sectors of focus for value investors are services and financial services: services represent 19.51% (19.19%) and financial services comprise 15.26% (14.65%) of the long (short) recommendations.

Panels C and D of table 4 present a summary of the financial data pertaining to the recommended securities used in the asset pricing tests. We find that the recommended investments are typically small with a slight tilt towards value. The median market capitalization (ME) is $397 million for long recommendations and the median book-to-market ratio (B/M) among long recommendations is 0.617. Based on the median Fama and French size and book-to-market breakpoints from 2000 to 2008, VIC recommendations are on average in the twentieth size percentile (indicating that 80% of stocks have higher market capitalizations than the average VIC stock) and the sixtieth book-to-market percentile (indicating that 40% of firms have book-to-market ratios higher than that of the average VIC recommended firm). This suggests that VIC recommendations are generally for small value stocks.

Among short recommendations the median market capitalization is $650 million, which is in the thirtieth size percentile—similar to long recommendations. However, the median book-to-market for short recommendations is much lower—0.342 (twenty-fifth percentile). The low median book-to-market suggests that when betting against a firm, VIC members focus on securities that would be considered "growth" on a book-to-market basis. With respect to profitability, long recommended firms are generally less profitable than the firms that are short recommendations. The median return on assets is 3.7% for long recommendations and 5.3% for short recommendations.

III. Performance Analysis

In this section we examine the performance of the recommendations made by VIC members. VIC recommendations typically state that their ideas should be considered "long-term" investments and not short term trades. To capture this notion of long-term performance, we perform detailed return calculations on horizons of one-, two-, and three-years. As Barber and Lyon (1997) argue, traditional event-time buy-and-hold abnormal returns (BHAR) "precisely measure investor experience" of buy-and-hold investors, the contingent most common in the value investing community.

An important potential issue when calculating long-run returns is the potential impact of firms that delist from public markets during the return horizon, whether due to acquisition, bankruptcy, or liquidation. We adjust returns to reflect delisting information using the methodology described in Beaver, McNichols, and Price (2007). We account for delisted firms in our analysis in a similar fashion to Lyon, Barber, Tsai (1999). If a firm is delisted (either a sample firm, or a control firm), we first incorporate the CRSP delisting return by compounding it with the cumulative return through the date of the delisting return. Then we assume the proceeds of the delisted firms are invested in the control-firm or benchmark portfolio. We also perform our analysis under the assumption that the delisted firms' proceeds are invested in the CRSP value-weighted index, and where delisted firms are eliminated from the database—the results are all very similar.

A. Control-firm BHAR

The control-firm event-time BHAR methodology we use follows that of Lyon, Barber and Tsai (1999). The model is represented as

$$\mathrm{AR_{it}} = \mathrm{R_{it}} - \mathrm{E(R_{it})}, \tag{1}$$

where $\mathrm{AR_{it}}$ is the BHAR to firm i in period t, $\mathrm{R_{it}}$ is the return generated by compounding successive monthly returns to firm i over period t and $\mathrm{E(R_{it})}$, is the compounded benchmark return in the same period.

Following the methodology of Speiss and Affleck-Graves (1995), we assign each sample firm a control-firm based on size (market value of equity) and book-to-market ratio. All firms in the CRSP/Compustat universe are considered potential matches. From the CRSP/Compustat universe we select as the control-firm that firm for which the sum of the absolute value of the percentage difference in size and the absolute value of the percentage difference in book-to-market ratio is minimized. We define size as the market value of equity on December 31 of the prior year and book-to-market ratio as book value of equity at the end of the last fiscal quarter of the prior calendar year divided by size.

We calculate the one-, two- and three-year BHARs to each recommendation using monthly CRSP data, following the advice of Brown and Warner (1985) who espouse the benefits of using monthly data rather than daily data. The event period return data begin on the first of the month following the date the recommendation was posted to the community. For example, if an idea is posted on January 15th, we start calculating monthly returns on February 1st. Because return data begins at the first of the month following the date of the recommendation, which leaves up to thirty days for VIC members to take positions, it is possible that the abnormal returns presented underestimate the true returns earned by VIC members, and may bias our tests in favor of the null hypothesis that fund managers have no stock-picking skills.

Table 5 presents summary statistics and results of the control-firm BHAR analysis. Abnormal returns to long recommendations are economically large and statistically significant. The value investors in our sample outperform control-firm benchmarks by 7.21% and 14.91% over the one-year and two-year periods following the VIC recommendation. However, the evidence from the short recommendation sample, although directionally correct, suggests we cannot reject the hypothesis that VIC members have no skill when shorting stocks. Nevertheless, because the short recommendation samples are small, we should not expect a rejection of the null hypothesis, since any long-term abnormal return test lacks statistical power in small samples (Ang and Zhang (2004)).

As a robustness test, we perform an alternate control-firm BHAR analysis. In these tests we further require that neither the size nor book-to-market ratio of the control-firm deviates from that of the sample firm by more than 10%. This ensures that sample firms examined are assigned a control-firm with very similar characteristics. The results from this analysis are very similar to those presented in table 5. We also perform the control-firm BHAR analysis after eliminating the top 1% and bottom 1% of observations to control for extreme outliers. The results are similar those presented in Table 5 (results not shown).

In addition to standard t-test p-values, we also present results in Table 5 from a sign test as per the recommendation by Ang and Zhang (2004), who conclude that the sign test coupled with a control-firm approach is well-specified and has the highest power for detecting long-term abnormal returns among competing long-term event study methods. This test also indicates statistically significant abnormal returns.

B. *Characteristic-based Benchmark-portfolio BHAR*

Savor and Lu (2009) suggest there are statistical issues with the control-firm BHAR methodology when the sample size is small and prone to outliers (as is the case with our sample of short recommendations). A remedy to this problem is the characteristics-based benchmark-portfolio BHAR approach, where the benchmark return is the return to a portfolio of stocks with characteristics similar to those of the sample stock. However the use of benchmark-portfolios reintroduces the skewness bias identified by Barber and Lyon (1997), which is mitigated by the control-firm BHAR approach. Therefore, in the analysis of statistical significance for the benchmark-portfolio BHAR approach we account for event-time skewness bias by using the bootstrapping method advocated by Lyon, Barber and Tsai (1999).

To construct the benchmark-portfolios we follow the characteristic-based benchmark methodology of Daniel, Grinblatt, Titman and Wermers (1997) (hereafter DGTW). We assign each stock in the CRSP universe to one of 125 portfolios containing securities with similar size, book-to-market and momentum characteristics. We then define the DGTW benchmark-portfolio abnormal return as the difference between the sample stock return and the DGTW benchmark-portfolio return, as in equation (1) above.

The results of this analysis are presented in Table 6. The results are consistent with the findings from the control-firm BHAR analysis. Using the DGTW benchmark-portfolio approach, we find that the investors in our sample generate statistically significant one-year BHARs of 9.52%, two-year BHARs of 19.03%, and three-year BHARs of 23.60% following the VIC recommendation.

For short recommendations, the DGTW benchmark-portfolio approach reaches the same conclusion as the control-firm BHAR approach: we cannot reject the null hypothesis of zero abnormal return once we adjust for skewness in the test statistics. However, unlike the control-firm BHAR analysis for short recommendations, the DGTW benchmark-portfolio BHARs are economically impressive: the matched sample one-year BHAR is 5.15%, two-year BHAR is 18.02% and three-year BHAR is 21.47%. Taken as a whole, the analysis of the short recommendations using the various BHAR approaches provides little statistical evidence that the investors in our sample are successful short sellers.

For robustness, we also perform the benchmark-portfolio BHAR analysis after eliminating the top 1% and bottom 1% of observations to control for extreme outliers. The results are similar to those presented in Table 6 (results not shown).

d. *Performance Analysis Discussion*

Regardless of a researcher's preference for BHAR methods, both methods presented in this study provide robust evidence that VIC members are successful in their long positions (see figure 1). The results of the various methods we use are statistically mixed with respect to VIC members' ability to successfully short stocks (see figure 2). Nevertheless, our overall conclusion from the evidence is that VIC members appear to have stock-picking skills. Figure 3 is an intuitive way to capture these results succinctly.

A potential critique of VIC members' recommendations is that these ideas are not implementable. This criticism is likely unwarranted. VIC is not set up so fund managers can showcase their ability to write research reports on opportunities that cannot be implemented—this would be a waste of the author's time and it would waste the membership's time. In fact, VIC has specific guidelines pertaining to the liquidity of investment recommendations submitted: "Small market capitalization ideas are fine, but as a general guideline, at least $250,000 worth of securities should trade on an average week. We understand that it is much more difficult to identify a compelling idea with $1billion of market capitalization, than one with $10mn of market capitalization and we take that into consideration when reviewing applications."

Another critique is that analyzing the full sample of VIC recommendations, without controlling for the "quality" of the recommendations may bias the results in favor of the null hypothesis that investors have no stock-picking skill. For example, if a member submits a really terrible idea because he was under time constraints, made mistakes in his analysis, or simply had no good ideas at the time, this idea may bias the results, even though the VIC member submitting the idea, and the broader VIC community can recognize the idea is no good. The next section explores this question in detail by looking at how ex-post long-term abnormal returns are related to the "quality" of an idea.

IV. The Relationship between VIC Ratings and Abnormal Returns

All VIC recommendations are not created equal. On September 14, 2009, the member "agape1095" posted a buy recommendation for Lehman Brothers, which was based on a serious mistake in the writer's analysis. The idea was given a rating of 1.3 by the VIC community—the worst rating in the history of VIC and more than five standard deviations below the mean for the entire sample. Moreover, on September 15, 2009, a VIC member posted a comment that the company had already entered bankruptcy. Agape1095 quickly replied, "I didn't know Lehman was already bankrupted when I posted this. And this report totally deserves the low rating."

The example above highlights why analyzing the full sample of VIC recommendations may misrepresent the skill of the majority of the investors in my sample. Although VIC membership is difficult to attain, even the best organizations can't completely screen out poor performers with certainty. To address this concern, I analyze how recommendations perform after controlling for the quality of the idea, as measured by the VIC community rating on individual investment thesis.

When investment recommendations are posted to VIC, members are given the opportunity to rate ideas on a scale of 1 (bad) to 10 (good). Ratings are recorded if five or more members rate the idea, and the rating period is open for two weeks. (Since 2007, which is when data is available on the time of rating, 60% of ratings were submitted within 72 hours of posting.) The club's guidance for ratings is that they should be objective and based purely on the quality of the investment thesis. Moreover, to encourage active participation, the club requires members to rate at least 20 ideas a year. The club also requests that extremely high (9 or 10) or extremely low (1 or 2) ratings be accompanied by some specific commentary about the investment thesis.

With the VIC ratings, I can perform additional tests to see if VIC members have an ability to pick stocks. In this analysis, I assume ratings approximate how favorably (or unfavorably) the VIC community believes the stock will perform in the future. To test whether VIC members can identify the best and worst recommendations within their universe of ideas, I estimate a simple model such that a linear relationship exists between abnormal returns and the VIC community rating. The model is represented as

$$BHAR_i = \delta_i + \lambda_i(Rating_i) + \varepsilon_i, \qquad (2)$$

where $BHAR_i$ is the abnormal return to stock i from *t*=2 to *t*=h (*h* is holding period), and $Rating_i$ is the VIC members' rating of the particular stock i. The dependent variable is calculated from *t*=2 to *t*=*h* to avoid an endogenous variable problem which may occur in a model that relates ratings with BHARs from *t*=1 to *t*=*h*. The endogenous variable problem may occur if an idea performs exceptionally well during the two-week rating period. For example, if stock X is recommended on June 20, 2008 and performs exceptionally well through July 3, 2008, members on July 3, 2008 may rate the idea extremely favorably (before the two-week rating period closes), not because they believe it will outperform in the future, but because it has performed well thus far.

In Table 12, I present coefficient estimates for the λ term in equation (2). I run regressions with the control-firm BHARs and benchmark portfolio BHARs as the dependent variable. The results suggest that VIC members have an ability to identify the best long recommendations posted to the website. Estimates for λ are positive and statistically significant across nearly all samples. I conclude from the evidence that VIC members are skilled at identifying the best and worst performing stocks within the universe of VIC recommendations.

The coefficients for the regressions performed on short recommendations also suggest the investors in my sample have an ability to discern between "good" short candidates and "bad" short candidates; however, this ability appears to be limited to a one-year horizon. The point estimates for $\hat{\lambda}$ are positive for the two- and three-year regressions; however, these estimates are not statistically significant so I cannot reject the hypothesis that VIC members cannot identify the best and worst short candidates over two- and three-year horizons.

I further analyze how ratings are related to abnormal returns by analyzing the abnormal returns associated with samples formed by rating quintiles. The results for the difference between the top and bottom rating quintile recommendations using control-firm BHAR analysis are in Table 13, and the results for the equivalent calendar-time portfolio regression analysis are in Table 14. The numbers from both tables provides strong evidence that VIC members have an ability to distinguish between "good" ideas and "bad" ideas. For example, the one-year BHAR abnormal returns associated with the top rating quintile are 21.69 percent, whereas the equivalent bottom quintile abnormal return is -.16 percent. Similarly, the average monthly alpha associated with the top rating quintile over a one-year horizon is 2.02 percent, compared to the bottom rating quintile which is –.27 percent. Figure 9 provides a visual presentation of these results.

The evidence clearly shows that VIC members have stock-picking skills. Not only is there evidence that the universe of VIC recommendations are successful on average, but there is strong evidence that members can distinguish ex-ante which stocks will outperform over the long term. This ability to distinguish between good and bad ideas is clearly a manifestation of stock-picking skill and provides ample evidence that skilled managers are in the investment management industry.

V. Conclusion

With our database, which is free from many of the biases found in databases analyzed by other researchers, we answer a basic economic question: (1) Are there *any* hedge fund managers that have stock-picking skill? Our answer is an emphatic, *yes.* We provide evidence that members of the Value Investors Club community can not only identify outperforming stocks on average, but that they can further distinguish among the best and worst of these outperforming stocks.

Our conclusion should not be surprising. The recommendations we analyze are well researched and required costly resources to develop. In equilibrium, skilled investors should be compensated for their efforts in accurately analyzing firms and driving assets to fundamental value (Grossman and Stiglitz (1980)).

In summary, this study brings into question the broader concepts of market efficiency in the asset markets; however, a key question remains concerning the magnitude of our findings. It is likely that the hedge fund managers we analyze control a relatively small portion of the total investment capital. Moreover, the evidence hints that the investors we analyze focus their efforts in small capitalization stocks and generally illiquid arbitrage situations. These asset classes may require additional risk factors which the asset pricing tests we utilize cannot account for. However, the economic significance of the large alpha point estimates in this study appear outsized relative to any reasonable compensation for systematic risk not accounted for with the current asset pricing models.

References

Ang, James and Shaojun Zhang, 2004, An evaluation of testing procedures for long horizon event studies, *Review of Quantitative Finance and Accounting* 23, 251-274.

Asness Cliff, Robert Krail, and John Liew, 2001, Do hedge funds hedge? Be cautious in analyzing monthly returns, *Journal of Portfolio Management* 28, 6-19.

Barber, Brad and John Lyon, 1997, Detecting long-run abnormal stock returns: The empirical power and specification of test statistics, *The Journal of Financial Economics* 43, 341-372.

Barber, Brad, Reuven Lehavy, Maureen McNichols and Brett Trueman. 2001. Can investors profit from the prophets? Security analyst recommendations and stock returns, *The Journal of Finance* 56, 531-563.

Barker, Robert, 2001, This message board may really light up, *Businessweek*, http://www.businessweek.com/archives/2001/b3722161.arc.htm, accessed October 30, 2008.

Beaver, William, Maureen McNichols, and Richard Price, 2007, Delisting returns and their effect on accounting-based market anomalies, *Journal of Accounting and Economics* 43, 341-368.

Berk, Jonathan and Richard Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269-1295.

Bollen, Nicolas and Veronica Pool, 2008, Conditional return smoothing in the hedge fund industry, *Journal of Financial and Quantitative Analysis* 43, 267–298.

Boyson, Nicole and Robert Mooradian, 2009, Hedge funds as shareholders activists from 1994-2005, Northeastern University working paper.

Brown, Stephen and Jerold Warner, 1985, Using daily stock returns: the case of event studies, *Journal of Financial Economics* 14, 3-31.

Cohen, Randolph, Christopher Polk, and Bernhard Silli, 2009, Best Ideas, Harvard University working paper.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristics based benchmarks, *The Journal of Finance* 52, 1257-1274.

Desai, Hemang and Prem C. Jain, 1995, An analysis of the recommendations of the superstar money managers at Barron's annual roundtable, *The Journal of Finance* 50, 1257-1273.

French, Kenneth, 1993, The cost of active investing, *The Journal of Finance* 63, 1537-1573.

Froot, Kenneth and Emil Dabora, 1999, How are stock prices affected by the location of the trade, *The Journal of Financial Economics* 53, 189-216.

Fung, William and David Hsieh, 2000, Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases, *Journal of Financial and Quantitative* Analysis 35, 291–307.

Fung, William and David Hsieh, 2001, The risk in hedge fund strategies: Theory and evidence from trend followers, *Review of Financial Studies* 14, 313–341.

Getmansky, Mila, Andrew Lo, and Igor Makarov, 2004, An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529–609.

Griffin, John and Jin Xu, 2009, How smart are the smart guys? A unique view from hedge fund stock holdings, *The Review of Financial Studies* 22, 2531-2570.

Grossman and Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen, 1995, Market underreaction to open market share repurchases, *Journal of Financial Economics* 39, 181-208.

Ljungqvist, Alexander, Christopher J. Malloy, and Felicia Marston, 2009, Rewriting history, *The Journal of Finance* 64, 1935-1960.

Lyon, John, Brad Barber, and Chih-Ling Tsai, 1999, Improved methods for tests of long-run abnormal stock returns, *The Journal of Finance* 54, 165-201.

Mitchell, Mark, Todd Pulvino, and Erik Stafford, 2001, Characteristics of risk and return in arbitrage, *The Journal of Finance* 56, 2135-2175.

Mitchell, Mark, Todd Pulvino, and Erik Stafford, 2002, Limited arbitrage in equity markets, *The Journal of Finance* 57, 551-584.

Munger, Charles, 2003, Academic economics: strengths and faults after considering interdisciplinary needs, Herb Kay Undergraduate Lecture, University of California, Santa Barbara Economics Department.

Savor, Nejat and Qi Lu, 2009, Do stock mergers create value for acquirers, *The Journal of Finance* 64, 1061-1098.

Seyhun, Nejat, 1988, The information content of aggregate insider trading, *Journal of Business* 61, 1-24.

Spiess, D.K., Affleck-Graves, J., 1995. Underperformance in long-run stock returns following seasoned equity offerings. *Journal of Financial Economics* 38, 243-267.

Table 1: Recommendation Summary Data

This table reports summary statistics for the sample of investment recommendations submitted to Valueinvestorsclub.com. The sample includes all recommendations shared with the VIC community from the time of the community's launch on January 1, 2000 through December 31, 2008. Panel A reports where assets are traded and the asset type recommended. Panel B reports the number of each long, short and long/short recommendation by the type of asset. Panel C reports the number of each long, short, and long/short recommendation by trading location.

Panel A: Asset type and trading location (n=3273)

Market	Common Stock	Bonds	Preferred Stock	Convertible Securities	Warrants	Options	Other	Total
US	2698	46	32	12	7	7	30	2832
Canada	156	1	2	0	0	0	2	161
UK/Europe	149	3	0	0	0	0	1	153
Japan	15	0	0	0	0	0	1	16
Hong Kong	19	0	0	0	0	0	0	19
Korea	14	0	0	0	0	0	0	14
Other	77	0	0	0	0	0	1	78
Total	3128	50	34	12	7	7	35	3273

Panel B: Recommendation by asset type (n=3273)

	Common Stock	Bonds	Preferred Stock	Convertible Securities	Warrants	Options	Other	Total
Long	2816	44	25	12	7	7	11	2922
Short	274	1	3	0	0	0	5	283
Long/Short	38	5	6	0	0	0	19	68
Total	2798	40	26	4	7	7	30	3273

Panel C: Recommendation and market location (n=3273)

	US	Canada	UK/ Europe	Japan	Hong Kong	Korea	Other	Total
Long	2508	158	139	15	17	13	72	2922
Short	273	0	7	0	0	0	3	283
Long/Short	51	3	7	1	2	1	3	68
Total	2832	161	153	16	19	14	78	3273

Table 2: Frequency of Criteria Cited as Basis for Recommendations

This table summarizes how frequently VIC members cited various criteria as the basis for their recommendations. Each recommendation is assigned at least one reason, and many ideas receive multiple criteria. Criteria were included if there were at least 10 recommendations that cited it as a unique criterion for investing in a particular asset.

N=3273

Criteria description	% of total
Intrinsic value undervaluation	86.83
Tangible asset undervaluation	23.62
Active open-market share repurchase program	11.73
Net operating loss assets	5.13
Recent restructuring, spinoff or spinoff potential	4.77
Insider buying	4.77
Undervaluation on a “sum-of-the-parts” basis	4.58
Involvement of activist investor	3.88
Lack of sell-side analyst coverage	2.69
Turnaround and/or recent bankruptcy	2.32
Liquidation potential	2.08
Complicated business or taxes creating investor confusion	1.89
Merger arbitrage situation	1.44
“Stub” arbitrage situation	1.34
Merger arbitrage trading opportunity	0.73
Pair-trade strategy	0.70

Table 3: Criteria Analysis

This table shows summary statistics for the sample of investment recommendations submitted to VIC between January 1, 2000 and December 31, 2008. Panel A highlights the top combinations of investment criteria used by value investors. Panel B reports the number of investment criteria used by investor recommendations submitted to VIC. (n=3273).

Panel A: Most common combinations

Rank	Criteria combination	# criteria	% of total
1	Intrinsic value	1540	47.05
2	Tangible assets; intrinsic value	299	9.14
3	Intrinsic value; share repurchase program	194	5.93
4	Tangible assets	150	4.58
5	Intrinsic value; net operating loss assets	70	2.14
6	Intrinsic value; restructuring, spinoff, or spinoff potential	67	2.05
7	Intrinsic value; insider buying	66	2.02
8	Tangible assets; intrinsic value; share repurchase program	61	1.86
9	Intrinsic value; sum of parts	57	1.74
10	Intrinsic value; activist investor involvement	38	1.16
Others		731	22.33

Panel B: # of criteria used

	#	% of total
1	1827	55.80%
2	1054	32.19%
3	325	9.93%
4	61	1.86%
5+	7	0.21%

Table 4: Recommendation Descriptive Statistics for Control-Firm Sample

This table reports summary statistics for the sample of VIC recommendations. The control-firm sample consists of all firms that have the necessary data to conduct the control-firm BHAR analysis. Panel A and B examine the distribution of investment recommendations using four-digit Standard Industry Classification (SIC) industries. Panel C and D show the characteristics of investment ideas. Panel E shows the frequency of recommendations by calendar year. B/M is the ratio of the LTM book value of equity to the market value of equity measured at the end of the month in which the investment is recommended. E/M is the ratio of LTM trailing earnings to the market value of equity measured at the end of the month in which the investment is recommended. ROA is the LTM return on assets. ME is the market value of equity measured at the end of the month in which the investment is recommended.

		Panel A: Industry representation for long recommendations		*Panel B: Industry representation short recommendations*	
Industry	SIC codes	Number of recommendations	Percent of sample	Number of recommendations	Percent of sample
Agriculture	< 1,000	8	0.48	4	2.02
Mining	1,000-1,499	69	4.13	4	2.02
Construction	1,500-1,999	22	1.32	4	2.02
Manufacturing	2,000-3,999	541	32.38	81	40.91
Transportation	4,000-4,999	170	10.17	10	5.05
Wholesale trade	5,000-5,199	63	3.77	8	4.04
Retail trade	5,200-5,999	199	11.91	20	10.10
Financial Services	6,000-6,999	255	15.26	29	14.65
Services	7,000-8,999	326	19.51	38	19.19
Other	> 9,000	11	0.66	0	0.00
No Data		7	0.42	0	0.00
Total		1671	100.0%	198	100.0%

Table 4: Recommendation Descriptive Statistics (continued)

Panel C: Long recommendation fundamental characteristics (n=1671)

	ME (millions)	B/M	E/M	ROA	ROE
Mean	4318	1.225	0.007	.029	0.011
25th Percentile	113	0.325	-0.006	-0.003	-0.010
Median	397	0.617	0.046	0.037	0.095
75th Percentile	1583	1.049	0.085	0.090	0.189

Panel D: Short recommendation fundamental characteristics (n=198)

	ME (millions)	B/M	E/M	ROA	ROE
Mean	2111	0.288	-0.100	0.087	0.400
25th Percentile	264	0.175	0.003	0.003	0.012
Median	650	0.342	0.037	0.053	0.121
75th Percentile	1738	0.668	0.067	0.108	0.221

Panel E: Time-series distribution of recommendations

Year	Long Recommendations	Short Recommendations
2000	95	1
2001	171	1
2002	181	10
2003	179	31
2004	190	25
2005	178	33
2006	196	30
2007	245	27
2008	236	40

Table 5: Control-Firm Buy-and-Hold Abnormal Returns

Returns to sample firms and control firms from January 1, 2000 to December 31, 2008. Control firms are selected by choosing the firm for which the sum of the absolute value of the percentage difference in size and the absolute value of the percentage difference in book-to-market ratio is minimized. The mean sample-firm returns and mean control-firm returns in panel B are returns to a short position in the security. P-values associated with a two-tailed paired t-test and a sign-test are presented. The sample consists of all firms that have the necessary data to conduct the control-firm BHAR analysis.

Panel A: Long recommendations

	N	Mean sample firm return	Mean control firm return	Difference (abnormal return)	P-value of t-test for difference	P-value of sign-test for difference
One-year	1429	17.28%	10.07%	7.21%	0.0015***	0.1010
Two-year	1152	43.34%	28.43%	14.91%	0.0003***	0.0087***
Three-year	945	72.34%	54.30%	18.04%	0.0066***	0.0007***

Panel B: Short recommendations

	N	Mean sample firm (short)	Mean control firm (short)	Difference (abnormal return)	P-value of t-test for difference	P-value of sign-test for difference
One-year	156	-4.16%	-7.05%	2.88%	0.6421	0.0924*
Two-year	128	-9.06%	-18.37%	9.32%	0.3275	0.2504
Three-year	97	-22.73%	-24.62%	1.90%	0.8784	0.1548

*, ** and *** denote two-tailed statistical significance at the 10%, 5% and 1% levels respectively.

Table 6: Benchmark-Portfolio Buy-and-Hold Abnormal Returns

Returns to sample firms and benchmark-portfolios from January 1, 2000 to December 31, 2008. Benchmark-portfolio abnormal returns are calculated by assigning each stock to one of 125 benchmark-portfolios based on size, book-to-market ratio, and momentum characteristics, then subtracting the benchmark-portfolio return from the sample firm return. Mean sample returns and mean benchmark-portfolio returns in panel B represent the return to a short position in the security or portfolio. P-values associated with a paired t-test and the Lyon, Barber, and Tsai (1999) bootstrapped skewness-adjusted t-statistics are also presented (1000 resamples of size=n/4). The sample consists of all firms that have the necessary data to conduct the benchmark-portfolio BHAR analysis.

Panel A: Long Recommendations

	n	Mean sample firm return	Mean benchmark-portfolio return	Difference (abnormal return)	P-value of paired t-test for difference	P-value for bootstrapped skewness-adjusted for difference
One-year	1327	17.11%	7.59%	9.52%	0.0000***	0.0000***
Two-year	988	45.02%	25.99%	19.03%	0.0000***	0.0000***
Three-year	777	74.39%	50.80%	23.60%	0.0000***	0.0013***

Panel B: Short Recommendations

	n	Mean sample firm return (short)	Mean sample firm return (short)	Difference (abnormal return)	P-value of paired t-test for difference	P-value for bootstrapped skewness-adjusted for difference
One-year	148	-2.02%	-7.17%	5.15%	0.0840*	0.4717
Two-year	115	-3.35%	-21.37%	18.02%	0.0014***	0.1877
Three-year	88	-12.74%	-34.21%	21.47%	0.0008**	0.4906

*, ** and *** denote two-tailed statistical significance at the 10%, 5% and 1% levels respectively.

Table 7: Predicting Matched-Sample Abnormal Returns with VIC Ratings

The regression model is given by $BHAR_i = \delta_i + \lambda_i(Rating_i) + \varepsilon_i$, where $BHAR_i$ is the cumulative abnormal return to stock i from t=2 to t=h (h is holding period), and $Rating_i$ is the VIC members' rating of a particular stock i. $\hat{\lambda}$, and $\overline{Rating}$ are sample estimates for the true parameters. VIC only reports a rating if five or more members rate a recommendation. The samples used in these regressions are the same one-, two-, and three-year samples used in the control-firm and benchmark-portfolio BHAR approaches. P-values associated with t-statistics are presented below the $\hat{\lambda}$ estimates (two-tailed).

	Control Firm BHAR			Benchmark Portfolio BHAR		
	One-year	Two-year	Three-year	One-year	Two-year	Three-year
Panel A: Long recommendations						
$\hat{\lambda}$	0.0808	0.0633	0.1925	0.0797	0.0793	0.1672
	0.0053***	0.1990	0.0157**	0.0002***	0.0520*	0.0070***
$\overline{Rating}$	5.10	5.11	5.11	5.10	5.11	5.13
Number of observations	1376	1123	928	1281	962	763
Panel B: Short recommendations						
$\hat{\lambda}$	.1233	0.0371	0.0980	0.1196	0.1004	0.0932
	0.1154	0.7470	0.5103	0.0323**	0.2642	0.4744
$\overline{Rating}$	5.33	5.36	5.26	5.32	5.36	5.31
Number of observations	152	124	95	144	111	86

*, ** and *** denote two-tailed statistical significance at the 10%, 5% and 1% levels respectively.

Table 8: Top and Bottom Rating Quintile Control-Firm Buy-and-Hold Abnormal Returns for Buy Recommendations

Returns to sample firms and control firms from January 1, 2000, to December 31, 2008. Control firms are selected by choosing the firm for which the sum of the absolute value of the percentage difference in size and the absolute value of the percentage difference in book-to-market ratio is minimized. The top (bottom) quintile for rating consists of the highest rated (lowest rated) 20% of the sample. P-values associated with a two-tailed paired t-test are presented. Test for difference between the top and bottom quintile p-values are calculated using a two-tailed paired t-test for difference assuming unequal variances.

Panel A: Top rating quintile

	n	Mean Sample Firm Return	Mean Control Firm Return	Difference (abnormal return)	P-value of t-test for difference	P-value of sign-test for difference	P-value of difference between top and bottom quintile
One-year	290	27.76%	6.07%	21.69%	0.0001***	0.0000***	.0017***
Two-year	255	46.59%	20.44%	26.15%	0.0028***	0.0024***	0.0536*
Three-year	221	86.86%	44.28%	42.58%	0.0054***	0.0054***	0.0061***

Panel B: Bottom rating quintile

	n	Mean sample firm return	Mean control firm return	Difference (abnormal return)	P-value of t-test for difference	P-value of sign-test for difference
One-year	254	8.56%	8.72%	-.16%	0.9736	0.8507
Two-year	202	32.26%	28.05%	4.21%	0.5737	0.8329
Three-year	168	46.25%	55.27%	-9.02%	0.4129	0.8170

*, ** and *** denote two-tailed statistical significance at the 10%, 5% and 1% levels respectively.

Figure 1

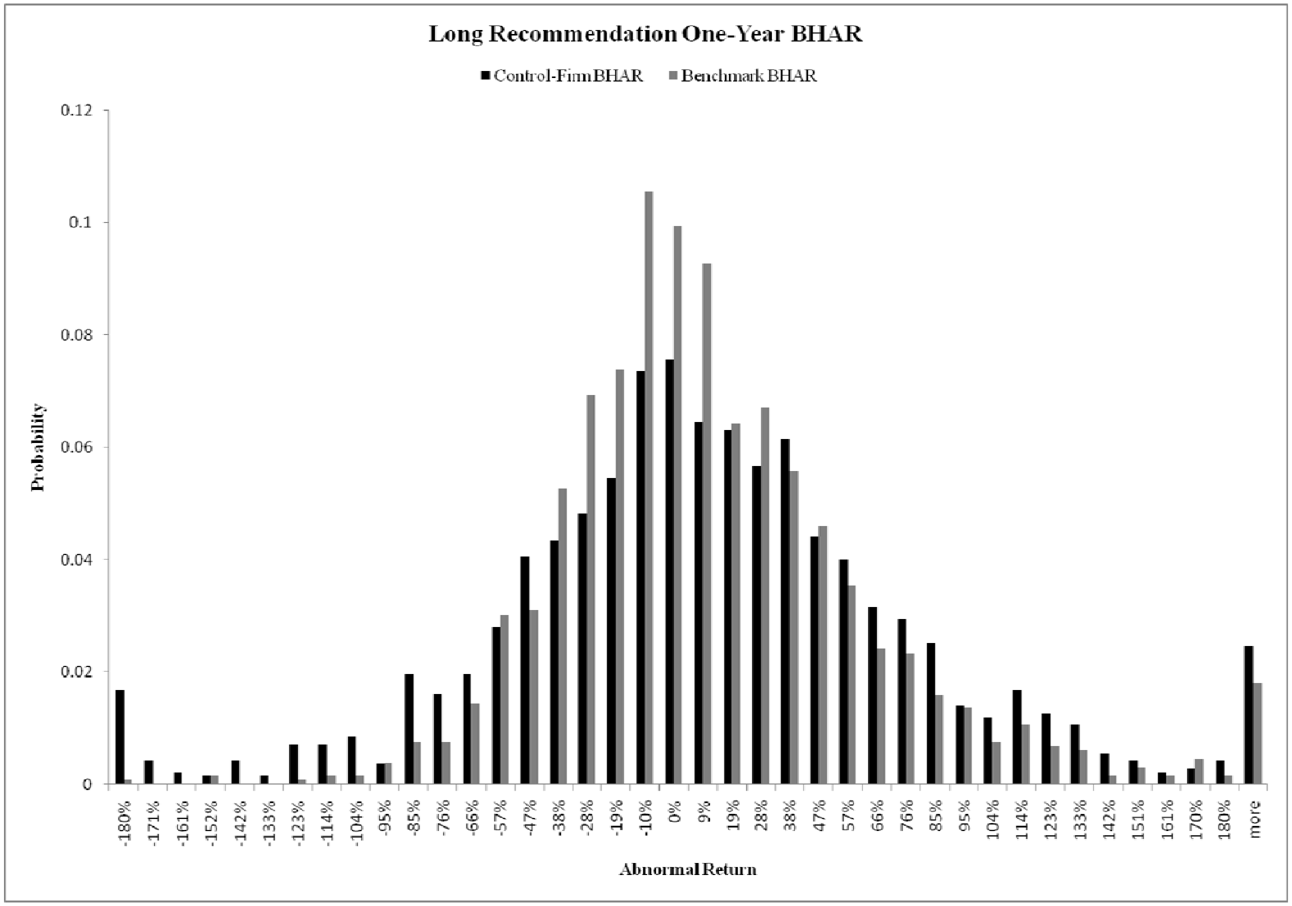

Figure 1: Long recommendation one-year BHAR. This figure represents the histogram of abnormal returns calculated from the control-firm and the benchmark-portfolio BHAR methodologies. The Y-axis represents the probability. The X-axis represents abnormal returns for long recommendations. The control-firm sample has 1,429 observations and the benchmark sample has 1,327 observations.

Figure 2

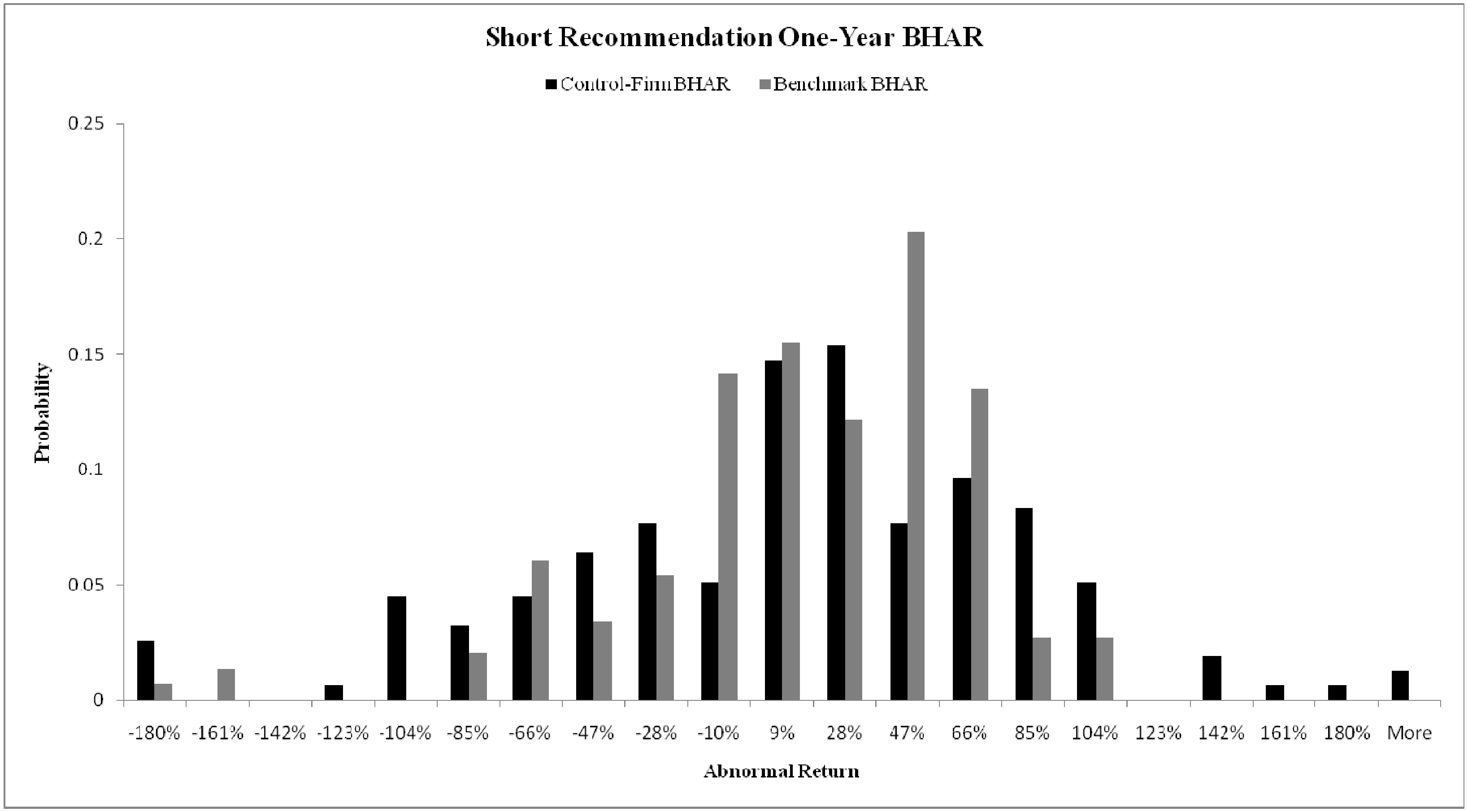

Figure 2: Short recommendation one-year BHAR. This figure represents the histogram of abnormal returns calculated from the control-firm and the benchmark-portfolio BHAR methodologies. The Y-axis represents the probability. The X-axis represents abnormal returns to a short position in short recommendations. The control-firm sample has 156 observations and the benchmark sample has 148 observations.

Figure 3

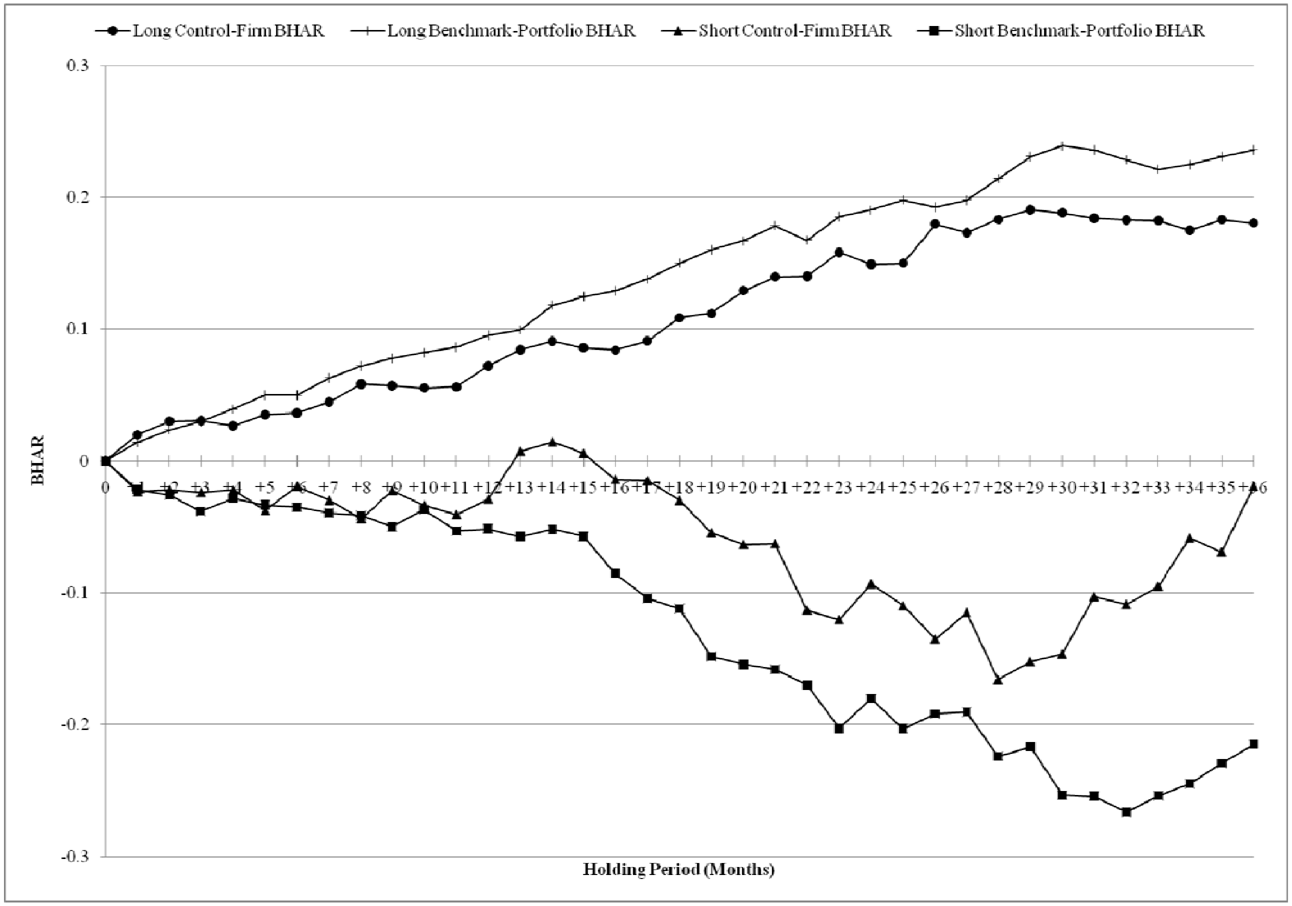

Figure 3: BHAR estimates for +1 to +36 months. This figure represents BHAR over time. The Y-axis represents the BHAR. The X-axis represents the holding period in months.

Figure 4

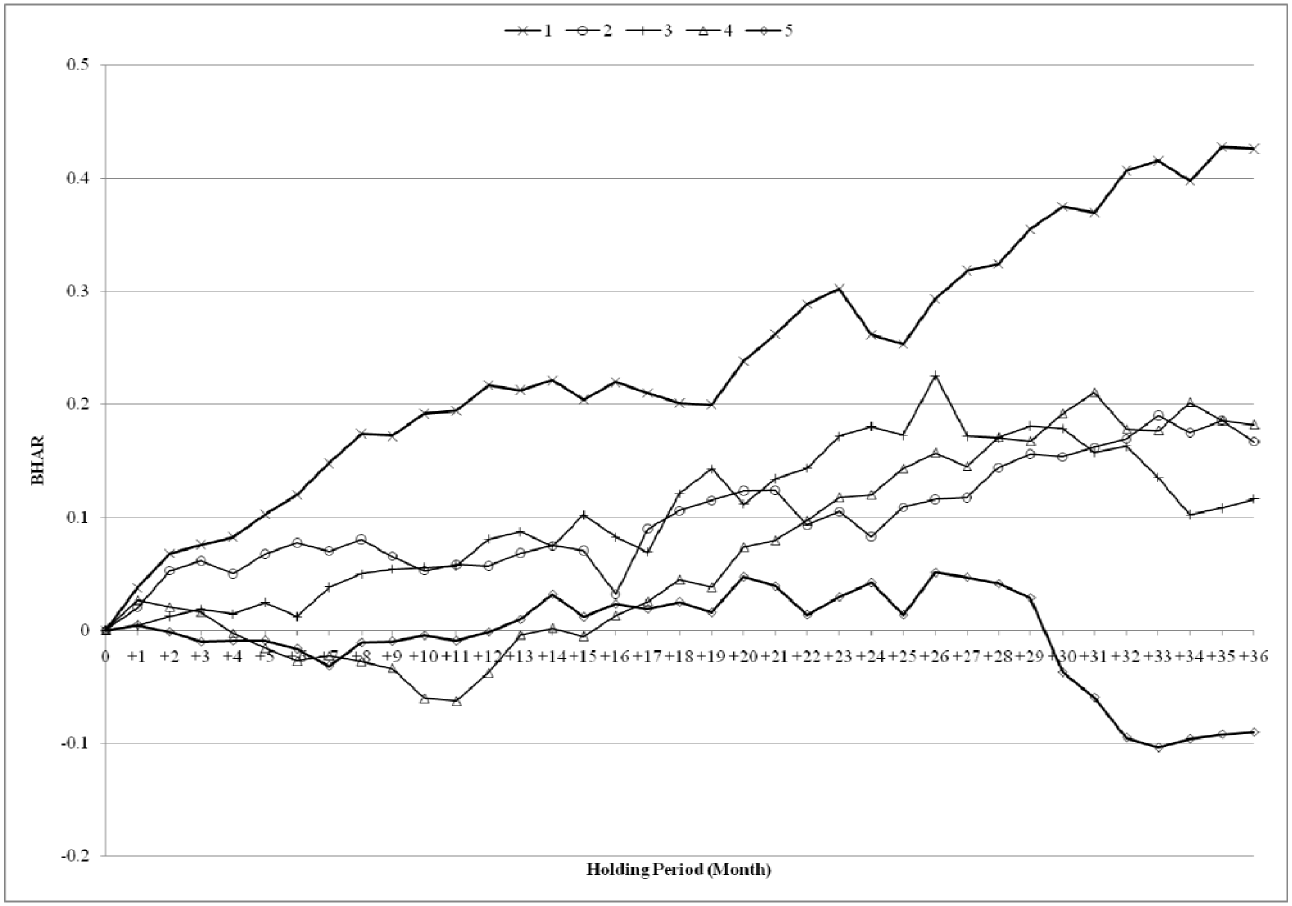

Figure 4: BHAR estimates for +1 to +36 months by rating (1=high, 5=low). This figure represents BHAR over time. The Y-axis represents the BHAR. The X-axis represents the holding period in months.