This article was translated by the author and reprinted from the November 2023 issue of the Securities Analysts Journal® with the permission of the Securities Analysts Association of Japan (SAAJ).

On the Persistence of Japanese Equity Mutual Fund Performance: An Analysis Using Stock Holding Data

Toshihiro Shimada, CMA

Abstract

This paper investigates the performance of Japanese equity mutual funds using stock holding data. We found that: 1) on average, style-adjusted stock selection skills are evident in the before-cost-deduction performance of these funds, 2) mid and small cap funds are the high performers, and 3) performance does persist. Thus, it is important to select funds exhibiting high style-adjusted stock selection skills and to manage their investment styles accurately in order to enhance investment portfolio performance stability.

Toshihiro Shimada, CMA

Mr. Shimada is a chief in the Investment Management Department of the Federation of National Public Service Personnel Mutual Aid Associations (KKR). He joined KKR in 2014, and after working at the Government Pension Investment Fund (GPIF), he has been in his current position since 2021. He graduated from Waseda University in 2014 and obtained an MBA from Hitotsubashi University Business School in 2023.

1. Introduction

The topics of whether active fund managers possess the ability to generate superior returns and whether investors should opt for passive or active funds have long been debated. Beginning with Jensen (1968), numerous studies have concluded that, on average, fund managers lack stock-picking abilities. However, as analytical methods and databases have advanced, research acknowledging the existence of superior funds under certain conditions has emerged, and this debate continues even now.

Recent trends in mutual funds indicate a continuous inflow of passive funds, which allow for low-cost operations. In the Japanese mutual fund market, the concept of long-term diversified investment has gained traction among individual investors, with the advantages of passive funds being actively highlighted. Additionally, in the management of pension funds and other institutional investors, the basic approach has been to focus on passive funds while allocating limited amounts to active funds.

Despite prevailing academic and practical skepticism towards active funds, there are criticisms of passive funds, such as concerns that increased investment in them could impair the market's price discovery function. Furthermore, recent events that have significantly impacted the equity market, such as Russia's military invasion of Ukraine and the tightening policies of central banks worldwide, suggest that relying solely on passive funds might make it difficult to achieve returns. This has led to a renewed interest in finding opportunities through active funds.

The purpose of this paper is to clarify whether active funds are valuable, with a focus on mutual funds that invest in Japanese equities. In addition to evaluating average performance, this paper will examine the factors that contribute to performance persistence and discuss methods for selecting mutual funds. A distinctive aspect of this paper is the use of mutual fund stock holding data for performance analysis. While analysis using stock holding data has been conducted in the United States since the 1990s, the accumulation of databases in Japan has not progressed, resulting in a limited number of studies. By utilizing stock holding data spanning approximately 15 years, this paper enables a more multifaceted analysis.

In Section 2, performance evaluation methods using mutual fund stock holding data will be discussed. Section 3 will describe the data, followed by analysis results in Sections 4 and 5. Finally, Section 6 will present the conclusions and suggest areas for further research.

2. Performance Evaluation Methodology

When evaluating the performance of mutual funds, methods such as Jensen (1968), the Fama and French (1993) three-factor model, and the Carhart (1997) four-factor model are commonly employed for estimating α. Analyzing mutual fund return data using these methods is relatively straightforward due to accessible and simple databases. However, as highlighted by Roll (1978), the choice of benchmark can significantly influence estimation results, posing a challenge known as the benchmark selection problem. To circumvent this issue, the use of stock holding data for analysis has gained traction.

In studies using actual stock holding data of mutual funds, Grinblatt and Titman (1993) argued that the covariance between the weights of held stocks and returns is expected to be positive for funds with stock selection abilities. Based on this notion, they introduced the GT measure. The GT measure for fund i in month t can be expressed as equation (1).

$$GT_{i,t} = \sum_{j=1}^{N} \left(w_{i,j,\,t-1} - w_{i,j,\,t-k-1}\right) R_{j,\,t} \tag{1}$$

where $w_{i,j,\,t-1}$ is the weight of stock j held by fund i at the end of month $t-1$, $w_{i,j,\,t-k-1}$ is the weight of stock j held by fund i at the end of month $t-k-1$, k is the lag (with $k = 12$ used in this paper for analysis), and $R_{j,t}$ is the return of stock j in month t. The GT measure can be viewed as the return of a strategy that goes long on the current portfolio and short on the portfolio from k months ago. Grinblatt and Titman (1993) demonstrated through the GT measure calculation that mutual funds achieve, on average, positive abnormal returns.

The GT measure, while circumventing the benchmark selection problem, does not fully consider anomalies such as size, value, and momentum effects. Indeed, Grinblatt *et al.* (1995) demonstrated that funds adopting momentum strategies exhibit higher GT measure. Therefore, Daniel *et al.* (1997) proposed a performance measure that controls for style using market capitalization, book-to-market ratio, and past one-year return. The performance measure reflecting fund i's style-adjusted stock selection ability in month t, hereinafter referred to as the characteristic selectivity measure (CS measure), can be expressed by equation (2).

$$CS_{i,t} = \sum_{j=1}^{N} w_{i,j,t-1} \left(R_{j,t} - R_t^{b_{j,t-1}}\right) \tag{2}$$

where $R_t^{b_{j,t-1}}$ is the return in month t of the benchmark used for stock j at the end of month $t-1$. Additionally, besides the CS measure, the measure due to style timing of fund i in month t, hereinafter referred to as the characteristic timing measure (CT measure), can be expressed in equation (3), and the measure due to average style bet, hereinafter referred to as the average style return measure (AS measure), can be expressed in equation (4).

$$CT_{i,t} = \sum_{j=1}^{N} \left(w_{i,j,\,t-1} R_t^{b_{j,t-1}} - w_{i,j,\,t-k-1} R_t^{b_{j,t-k-1}} \right) \tag{3}$$

$$AS_{i,t} = \sum_{j=1}^{N} w_{i,j,\,t-k-1} R_t^{b_{j,t-k-1}} \tag{4}$$

In this approach, it is possible to decompose the fund's return into the CS, CT, and AS measures.[1] In this paper, we use a lag of $k = 12$, similar to the GT measure, and employ 125 benchmarks sorted into quintiles based on market capitalization, book-to-market ratio, and past one-year return, following the principles outlined by Daniel *et al.* (1997) and taking into account characteristics specific to Japanese companies, such as the fiscal month, as discussed by Kubota and Takehara (2007).[2]

Daniel *et al.* (1997) revealed through the above approach that while mutual funds on average possess style-adjusted stock selection ability, the annualized CS measure is less than 1%, only about the same as management fees, and there is no evidence of style timing ability. Analysis using stock holding data is challenging due to data collection difficulties, hence there have been few studies conducted in Japan. Among them, Asakura and Uno (2004), using Japanese pension equity fund stock holding data, revealed tendencies for performance to decrease with stronger momentum, though statistically insignificant.

3. Data

For our analysis, we used Morningstar Direct, provided by Ibbotson Associates Japan, to select general-type actively managed domiciled open-end mutual funds that invest in Japanese equities.[3] Of these, 698 funds that existed from August 2007 to December 2021, when stock holding data became available, were included in the analysis. The stock holding data is updated once or twice a year in principle, as it contains data as of the end of each fund's fiscal period.[4] Fund return data, etc., were obtained from data on domiciled open-end mutual funds

---

[1] The sum of each measure can be represented as follows, allowing for the decomposition of the gross return of fund i in month t into three measures.

$$\begin{aligned} CS_{i,t} + CT_{i,t} + AS_{i,t} &= \sum_{j=1}^{N} \left(w_{i,j,t-1} \left(R_{j,t} - R_t^{b_{j,t-1}} \right) + \left(w_{i,j,\,t-1} R_t^{b_{j,t-1}} - w_{i,j,\,t-k-1} R_t^{b_{j,t-k-1}} \right) + w_{i,j,\,t-k-1} R_t^{b_{j,t-k-1}} \right) \\ &= \sum_{j=1}^{N} w_{i,j,\,t-1} R_{j,\,t} \end{aligned}$$

[2] Refer to Appendix A for details on the benchmark construction method.

[3] We excluded funds dedicated to defined contribution plans, fund wraps, SMAs, and ETFs. Additionally, we excluded funds with limited additions, early redemption conditions, or currency selection.

[4] Thus, analysis using holding data has the problem of not being able to accurately reflect the actual status of the fund due to the infrequent data updates. Our analysis assumes that the fund is rebalanced monthly to the portfolio weight calculated based on the most recent stock holding data at the end of each month.

provided by Financial Data Solutions (hereinafter "FDS"). This dataset includes daily information such as the rate of return for reinvested dividends before tax and after deduction of trust fees, net asset value, the amount of funds set up, and the amount of funds canceled. These daily data were converted to monthly data for each fund and used for the analysis in this paper.

The summary statistics for the Japanese equity mutual fund dataset are shown in Table 1. The excess return versus TOPIX is positive on average, but this return is after the deduction of trust fees and before the deduction of purchase commissions, so the excess return versus TOPIX would be negative if purchase commissions for the average holding period of the mutual funds were taken into account.[5] The number of funds was extremely small at the beginning, but the number of funds from which stock holding data could be obtained continued to increase during 2008, rising from 68 funds in January 2008 to 399 funds in December 2008. Since then, a sufficient number of funds has been included.[6]

Table 1

Summary Statistics for Japanese Equity Mutual Fund Dataset, August 2007–December 2021

Variable	Observations	Mean	SD	Min	P25	P50	P75	Max
Excess return vs. TOPIX (%/month)	698	0.057	0.389	-3.426	-0.115	0.023	0.197	1.643
Investment period (month)	698	98.155	55.793	1	50	91.5	161	173
NAV (¥100 million)	698	61.438	165.395	0.010	6.623	19.491	52.832	3,373.946
Cash flow (¥100 million/month)	698	-0.637	1.785	-13.925	-0.586	-0.167	-0.032	5.607
Trust fee (%/year)	698	1.558	0.300	0.110	1.430	1.650	1.716	2.607
Purchase commission (%)	698	2.694	0.760	0	3	3	3	5
Number of funds	173	396.023	83.880	2	402	423	431	440

Source: Prepared by the author (same as below).
Note: "Excess return vs. TOPIX" is the monthly excess return versus TOPIX (including dividends). "Investment period" is the number of months each fund was included in the analysis. "NAV" is the average net asset value during the investment period. "Cash flow" is the average cash flow during the investment period. "Trust fee" is the trust fee rate including tax. "Purchase commission" is the maximum purchase commissions. "Number of funds" is the number of funds included in the analysis for each month.

Furthermore, using the stock holding data, we calculated style indices based on market capitalization, book-to-market ratio, and past one-year return. Additionally, we counted the number of holdings.[7] First, examining the distribution of style indices, we observe that the funds analyzed tend to favor large-cap, growth, and momentum stocks. In particular, there is a pronounced bias towards large-cap stocks. This result is expected, considering that small-cap stocks are often excluded from the investment universe due to liquidity considerations

[5] Nishiuchi *et al.* (2019) measured the average holding period of mutual funds as 2.774 years using data on the number of beneficial ownership units.
[6] The minimum number of funds in Table 1 is 2, which corresponds to August 2007, the initial month of the analysis period.
[7] Refer to Appendix B for the method used to calculate the style indices and the distributions of both the style indices and the number of holdings.

and that many funds are managed with a focus on market capitalization-weighted indices such as TOPIX. For other style indices, there is no more pronounced bias than for size. A comparison of the average of all funds and TOPIX shows a tendency toward small-cap, growth, and momentum relative to TOPIX, although the divergence is limited. Next, examining the distribution of the number of holdings reveals that, on average, funds hold just under 100 stocks, with the majority holding 150 or fewer stocks. Considering the total number of stocks in the market, the number of holdings by each fund is extremely limited.

4. Results

Table 2 shows the performance measures for Japanese equity mutual funds. In addition to the analysis results for all funds, this section also shows the results for each category classified according to Morningstar Direct's definition.[8] Note that the gross returns of the funds, obtained by summing the CS, CT, and AS measures, are only hypothetical returns calculated using the stock holding data, and differ from actual returns. Therefore, before discussing the analysis results, we checked the deviation between hypothetical returns and actual returns, and determined that the hypothetical returns generally capture the actual status of the fund's operations.[9]

First, we review the results of the four-factor model using net return data.[10] The α_FFC4 for all funds in Table 2 is –0.152%, although not statistically significant. Given that the data used in the analysis is before the deduction of purchase commissions, it is unlikely that, on average, the ability to select stocks is sufficient to outweigh the cost. By category, the alpha for growth funds is relatively high, but none of the results are statistically significant.

We then review the average performance measure for all funds calculated from the stock holding data.[11・12] Looking at the GT measure, which shows stock selection ability before style adjustment in Table 2, it is a negative value.[13] However, the CS measure, which shows stock

---

[8] In principle, each fund was classified into one of six categories (2 × 3 = 6): two size categories (large and mid-small) and three style categories (blend, value, and growth). Additionally, the funds for analysis were classified into one of a total of seven categories, including specific region/sector.

[9] Refer to Appendix C.

[10] The analysis in this paper used factor returns provided by FDS. Refer to Appendix D for details.

[11] Cash held by the funds and stocks that have not yet been listed were excluded from the analysis, as they cannot be used to calculate the respective performance measures. Therefore, the analysis was conducted after standardizing the weight of each stock so that the total weight of stocks subject to analysis at each point in time for each fund was 100% (the same applied to the subsequent analysis).

[12] We examined the impact of the benchmark construction method on the CS, CT, and AS measures. Refer to Appendix E for results.

[13] The first year of the analysis period, 2009, marked a reversal from the global financial crisis, and the return on the momentum factor was very negative. Since the funds in the analysis, on average, adopted momentum-oriented strategies, it can be inferred that this contributed to the deterioration of the GT measure.

selection ability after style adjustment, is 1.364%, which is significantly positive at the 1% level, suggesting that on average there is stock selection ability after style adjustment. However, considering the levels of trust fees and purchase commissions, the performance of the CS measure is offset by these costs, indicating that mutual fund investors are not able to benefit from the fund manager's stock selection ability. Many analyses using stock holding data, including Daniel *et al.* (1997), suggest that fund managers possess excellent skills, but they also point out that the performance achieved is merely equivalent to the costs, such as trust fees. The analysis in this paper obtained similar results.

Table 2

Performance Measures, January 2009–December 2021

Category	Number of funds	α_FFC4 (%/year)	GT (%/year)	CS (%/year)	CT (%/year)	AS (%/year)
Large blend funds	125	-0.721	-0.456	0.928**	0.226	10.142
		(-1.542)	(-0.786)	(2.003)	(0.588)	
Large value funds	27	-0.815	-0.127	1.078**	0.246	10.341
		(-1.184)	(-0.301)	(2.196)	(0.557)	
Large growth funds	61	0.375	-1.105	1.295**	0.336	10.816
		(0.546)	(-1.219)	(2.166)	(0.851)	
Mid-small blend funds	44	-0.382	-0.384	1.360***	0.289	10.990
		(-0.591)	(-0.715)	(2.756)	(1.009)	
Mid-small value funds	11	-0.274	0.361	1.385***	-0.024	11.020
		(-0.468)	(1.145)	(3.156)	(-0.089)	
Mid-small growth funds	84	1.111	-1.007	2.427***	-0.020	12.163
		(0.773)	(-1.016)	(2.655)	(-0.070)	
Specific region/sector funds	34	-0.773	-0.652	1.074*	0.004	10.745
		(-1.145)	(-1.274)	(1.675)	(0.014)	
All funds	385	-0.152	-0.711	1.364***	0.149	10.810
		(-0.253)	(-1.116)	(2.752)	(0.524)	

Note: α_FFC4 is estimated using Carhart's (1997) four-factor model, GT is estimated using equation (1), CS using equation (2), CT using equation (3), and AS using equation (4). All values are annualized. The analysis period for large value funds is from February 2009 to December 2021, as there were no applicable funds in January 2009. Figures in parentheses represent *t*-values, with * indicating statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

In addition, we will also confirm the CT and AS measures to see whether value is being added through style timing and style betting. The CT measure for all funds is slightly positive but not statistically significant. On average, fund managers do not possess the ability to perform style timing, a point also noted by Daniel *et al.* (1997). The gross return of 11.397% is higher than TOPIX, which returned 8.996% over the same period. The level of excess return exceeds the sum of the CS and CT measures, which can be attributed to the AS measure, representing an average style bet. If the style bet is intended to obtain excess returns, it can be attributed to the fund manager's skill. However, if the style bet results from the fund's concept or investment style, it cannot be concluded that the fund manager has the ability to make effective style bets. The analysis in this paper does not extend to verifying whether the fund manager can create added value through intentional style bets.

Finally, we mention the performance measures for each category. The GT measure is negative except for mid-small value funds, but we cannot confirm its statistical significance in any category. The CS measure is significant at the 1% level for each of the mid-small categories, suggesting high performance of mid-small funds. In particular, mid-small growth funds have achieved performance approximately 1% per year higher than other categories. The CT measure tends to be relatively higher for large funds, but there are no significant differences among the categories, and the results are not significant for all categories. The AS measure tends to be generally higher for mid-small funds, due in part to the higher return on the size factor over the analysis period. Among them, the level of mid-small growth funds is high and can be said to have earned their returns from average style bets as well as style-adjusted stock selection ability, but as noted earlier, this is not necessarily due to the superior ability of fund managers.

Although Grinblatt and Titman (1993) and Daniel *et al.* (1997) have revealed high performance of growth funds, such trend was not clearly observed in the analysis performed for this paper. On the other hand, it can be pointed out that the high performance of mid-small funds may be a Japan-specific trend. One possible factor could be that a large number of stocks with smaller market capitalization are listed compared to the U.S. stock market. However, the analysis in this paper does not clarify whether there are many talented managers of mid-small funds or whether the mid-small market has inefficiencies that make it easier for them to earn returns. It will also be necessary to closely monitor whether the recent market reclassification by the Tokyo Stock Exchange and the growing interest in cost-of-capital-conscious management will have any impact on the performance of active funds.

5. Performance Persistence

What factors should be considered when investing in funds expected to perform well in the future? The simplest approach is to use past performance as a reference. It is common for manager structures at pension funds to consider past performance to some extent when selecting funds. In this section, we will examine whether the funds analyzed in this paper exhibit performance persistence and discuss fund structures that ensure stable returns.

Regarding performance persistence, Jensen (1969) has recognized a positive correlation between past and future performance. Grinblatt and Titman (1993) reported the persistence of the GT measure in their analysis using stock holding data. In an analysis of funds investing in Japanese equities, Uno (2002) revealed some persistence in performance, and Shikata (2012) found persistence in style-adjusted alpha for funds with consistently high past performance.

First, to confirm performance persistence in the funds analyzed in this paper, we created decile portfolios sorted by actual returns over the past one, three, and five years. Each decile portfolio was assumed to be rebalanced monthly according to the actual returns up to the end of the previous month and invested equally in the funds within that portfolio. The returns at the time of sorting and the subsequent returns obtained by this procedure are shown in Table 3. Table 3 indicates a certain degree of performance persistence. In addition to the observation that higher returns at the time of sorting correspond to generally higher post-sort returns, the difference in post-sort returns between the tenth decile with the highest historical returns and the first decile with the lowest historical returns is significant. The number of years of past returns used for sorting also shows that the difference in post-sort returns between the tenth decile and the first decile increases as longer past returns are considered. This suggests that it is preferable to refer to longer-term historical returns when selecting funds.

Table 3

Decile Portfolio Returns, January 2013–December 2021

	Sort by past 1-year returns		Sort by past 3-year returns		Sort by past 5-year returns	
Decile portfolios	At sorting (net, %/year)	Post-sorting (net, %/year)	At sorting (net, %/year)	Post-sorting (net, %/year)	At sorting (net, %/year)	Post-sorting (net, %/year)
1 (low)	3.449	12.533	4.892	11.875	5.166	11.187
1/2	6.187	12.668	6.316	12.032	6.259	11.694
3/4	12.112	12.598	9.266	11.874	8.538	12.135
5/6	15.362	12.399	11.003	12.282	9.921	12.402
7/8	19.192	13.046	13.402	13.533	12.050	13.928
9/10	29.463	17.046	20.135	17.691	18.068	17.594
10 (high)	34.864	18.866	23.645	19.618	21.193	19.771
10-1 spread	31.415	6.333*	18.752	7.743**	16.026	8.583**
		(1.668)		(2.067)		(2.130)

Note: Figures in parentheses represent *t*-values, with * indicating statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

What factors contribute to performance persistence? Figure 1 shows the results of decomposing the cumulative excess returns of the tenth decile relative to the first decile into CS, CT, AS measures, and others. This shows that the return difference is stable by the CS measure, regardless of the number of years used for sorting. The results of sorting by past three- and five-year returns show that the AS measure also produces return differences, but they are not as stable as the CS measure. This indicates that performance persistence is mainly caused by the stock-picking ability after adjusting for style, while returns from style bets, which are affected by market conditions, are not necessarily a factor in performance persistence. Uno (2002) and Shikata (2012) have pointed out the persistence of style-adjusted alpha for Japanese equity funds using factor models, and the analysis in this paper supports these findings. However, there are also several return differences that cannot be decomposed into CS, CT, or AS measures. These differences arise from the divergence between actual and hypothetical returns. The actual return in the tenth decile is higher than the gross hypothetical return, which could be hypothesized as a result of transactions not captured by the stock holding data. However, this factor has not been clearly examined. It remains to be seen whether other factors besides the CS measure contribute to stable performance.

Figure 1
Cumulative Excess Returns of the Tenth Decile Minus the First Decile, January 2013–December 2021

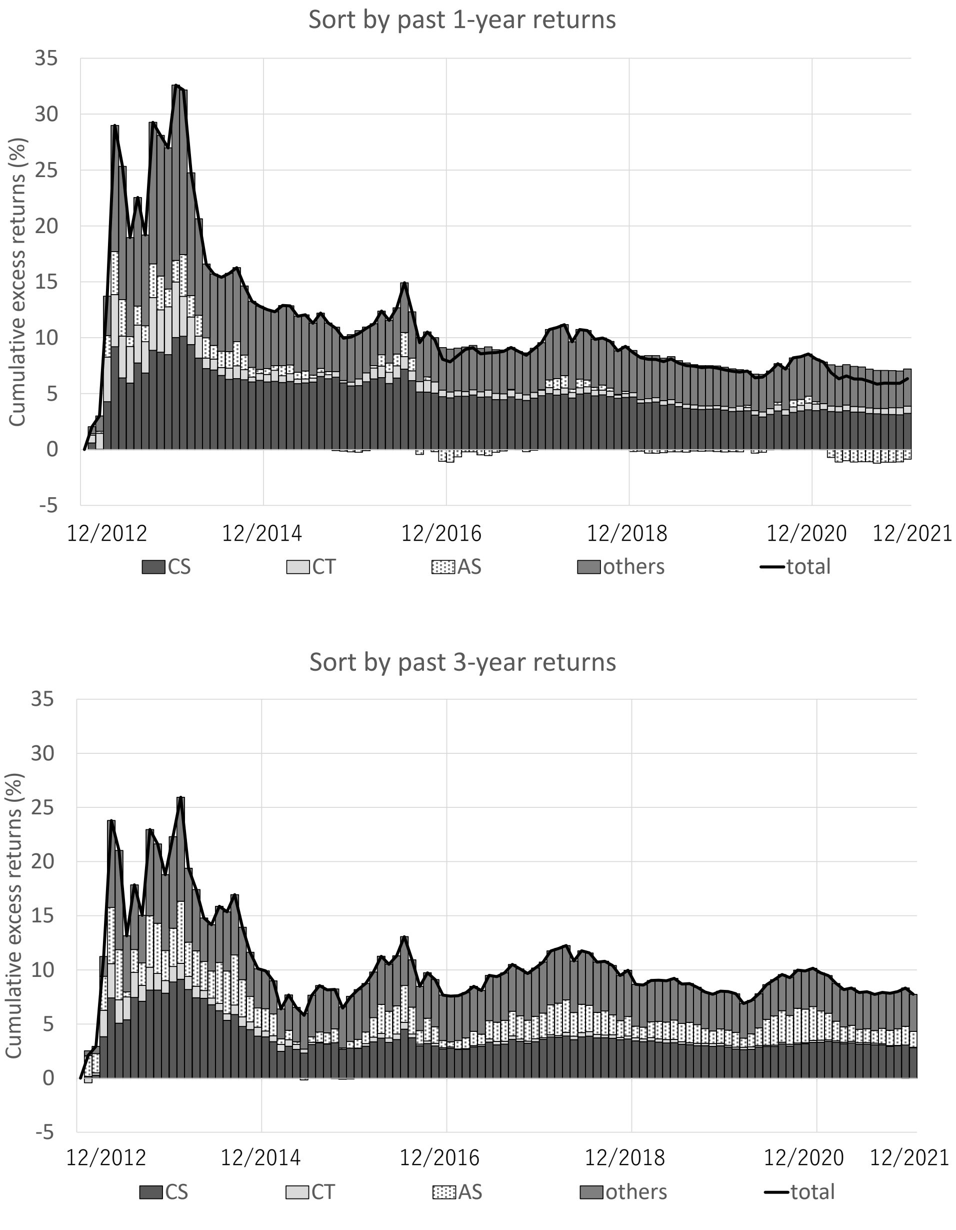
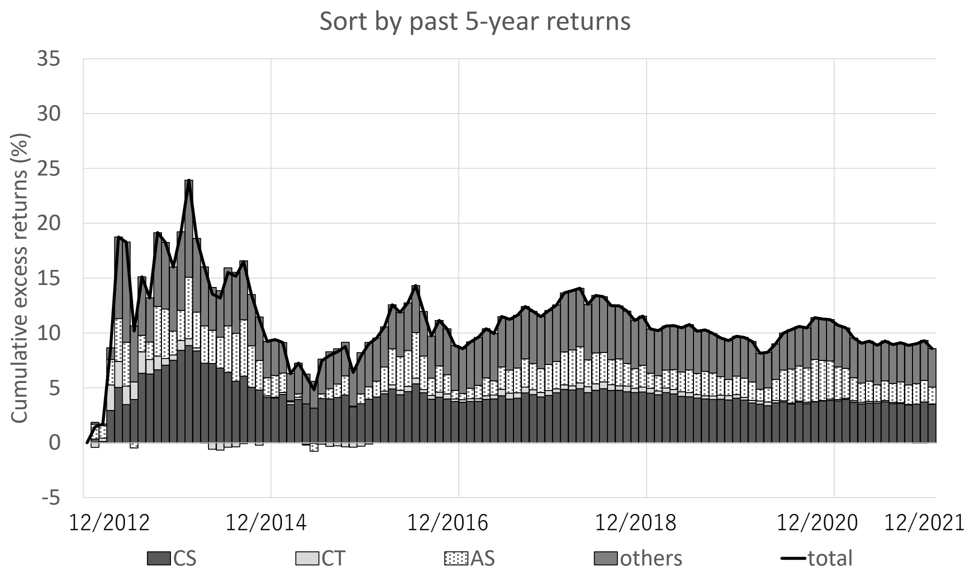


Figure 1 (Continued)



Note: Cumulative excess returns since January 2014 are annualized.

Although not all factors are identified, the results indicate that performance persistence is also observed in the funds analyzed in this paper. While investing in funds with high past returns is expected to contribute to higher future returns, it is also possible that the fund composition may be skewed toward a particular style. Depending on market conditions, these high returns may not be sustained. To achieve stable excess returns relative to the market benchmark, it is necessary to accurately manage the overall style of the portfolio. Additionally, an analysis of persistence in each performance measure shows that the CS measure exhibits persistence.[14] Therefore, it would be effective to reference past CS measures to some extent and select funds with high style-adjusted stock-picking ability.

6. Conclusion and Further Research

Using stock holding data, this paper confirmed the average performance of Japanese equity mutual funds and examined the factors that contributed to their performance persistence. First, the mutual funds analyzed were, on average, found to have stock-picking ability before deducting expenses. Although fund managers exhibited a certain degree of superior ability, this stock-picking ability disappeared when trust fees and purchase commissions were taken into account, indicating that mutual fund investors could not benefit from it. Additionally, style timing ability was not observed, and mid-small funds achieved relatively high performance.

[14] Refer to Appendix F.

Furthermore, an analysis of the factors contributing to performance persistence showed a positive relationship between past and future performance. In particular, stock selection ability after style adjustment tended to remain stable. In order to obtain stable excess returns relative to market benchmarks such as TOPIX, it is important to select funds with high style-adjusted stock selection ability and to accurately manage the overall portfolio's style. Additionally, the fact that subsequent returns were higher when sorted by long-term returns and that persistence was observed in style-adjusted stock selection ability itself may be useful information for fund selection.

Although the market environment has undergone significant changes in recent years, such as frequent increases in stock market volatility and shifts from growth to value markets, we were able to demonstrate the possibility of increasing expected returns by devising certain fund compositions. While it is significant that we found a certain degree of value in investing in active funds, there are limitations to analyses based on stock holding data. In principle, the stock holding data used in this paper is updated only once or twice a year, making it impossible to capture transactions during the intervening periods. This limitation hinders a full examination of the data. From the perspective of promoting information disclosure to investors, it is hoped that an environment will be created where the frequency of stock holding disclosures is increased.

Even with such data limitations, our future task is to propose a fund structure that can generate stable excess returns. For example, more in-depth analysis is required on how to select funds with high style-adjusted stock selection ability and on portfolio style management methods based on an understanding of each fund's style characteristics and the degree of style drift.

References

Asakura, Nobuhito and Yoko Uno, 2004, "Investigation of Manager Trading and Herding - Using Composition Data of Equity Portfolios," *Securities Analysts Journal*, 42(6), pp. 109–125 (in Japanese).

Uno, Yoko, 2002, "Return Persistence in Pension Equity Funds," *Securities Analysts Journal*, 40(11), pp. 96–110 (in Japanese).

Kubota, Keiichi and Hitoshi Takehara, 2007, "Revalidation of Effectiveness of Fama-French Factor Model," *Gendai Finance*, 22, pp. 3–23 (in Japanese).

Shikata, Takehiko, 2012, "Performance Evaluation Based on Relative Ranking," *Securities Analysts Journal*, 50(5), pp. 68–78 (in Japanese).

Nishiuchi, Sho, Toshiki Honda and Daisuke Miyakawa, 2019, "Performance and Size of Mutual Funds," *Shintaku Kenkyu Syoureikin Ronsyu* , 40, pp. 104–114 (in Japanese).

Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52(1), pp. 57–82.

Daniel, K., M. Grinblatt, S. Titman and R. Wermers, 1997, "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks," *Journal of Finance* 52(3), pp. 1035–1058.

Fama, E. F. and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33(1), pp. 3–56.

Grinblatt, M. and S. Titman, 1993, "Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns," *Journal of Business* 66(1), pp. 47–68.

Grinblatt, M., S. Titman and R. Wermers, 1995, "Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior," *American Economic Review* 85(5), pp. 1088–1105.

Jensen, M. C., 1968, "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance* 23(2), pp. 389–416.

Jensen, M. C., 1969, "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business* 42(2), pp. 167–247.

Roll, R., 1978, "Ambiguity when Performance is Measured by the Securities Market Line," *Journal of Finance* 33(4), pp. 1051–1069.

Appendix A

Benchmark Construction Method Based on the Approach of Daniel *et al.* (1997)

This appendix describes the method used in this paper to construct benchmarks for calculating CS, CT, and AS measures.

First, the universe of component stocks included in the benchmarks consists of all common stocks listed in Japan, while the sorting universe is limited to stocks listed on the Tokyo Stock Exchange First Section. The sorting timing is set for the last business day of August each year, and stocks that are newly listed in September or later are not included until the next sorting period (the last business day of August the following year). Stocks with no or negative actual equity capital and stocks with no past one-year returns are excluded from both the component universe and the sorting universe.

At the sorting time on the last business day of August each year, the component universe is first divided into five groups based on the market capitalization of common stocks at the end of August. The quintile points are determined by the stocks in the component universe that are listed on the TSE First Section. Thus, the same number of stocks are allocated to each quintile within the TSE First Section, but the number of stocks in each quintile is not equal across the component universe, including stocks not listed on the TSE First Section.[15]

Next, each quintile portfolio is further divided into five portfolios based on the book-to-market ratio, resulting in a total of 25 portfolios. The book value used for the book-to-market ratio is the equity capital for the most recent fiscal year available as of the end of August (with consolidated financial statements being prioritized), and the market value is the market capitalization of common stock at the end of August, which is used in the initial division. Finally, each of these 25 portfolios is further divided into five groups based on past one-year returns, resulting in a total of 125 benchmarks. The past one-year returns are calculated using the dividend-inclusive rate of return from August of the previous year to July of the current year (i.e., one year up to one month before the sort timing).

For each benchmark created using this method, monthly returns are calculated based on the market capitalization-weighted dividend-inclusive rate of return of the common stocks within each benchmark. These returns are then utilized in the analysis presented in this paper. For example, the CS measure is obtained by calculating the difference between the return in month t of the stocks held by the fund at the end of month $t-1$ and the return of the benchmark that includes those stocks in month t. This difference is then weighted by the proportion of each stock held to derive the fund's CS measure for month t.

---

[15] Strictly speaking, unless the number of stocks in the sort universe is a multiple of five, the number of stocks listed on the TSE First Section in each quintile will not be exactly equal.

Appendix B

Calculation Method of Style Indices and Distribution of Style Indices and Number of Holdings

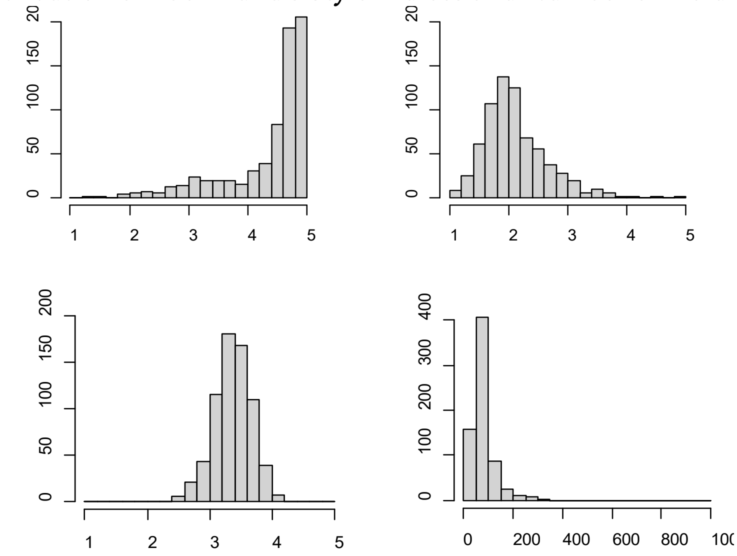
This appendix describes how to calculate style indices using stock holding data, based on market capitalization, book-to-market ratio, and past one-year return.

First, at the end of August of each year, all common stocks listed in Japan are classified into quintiles based on each stock's market capitalization, book-to-market ratio, and past one-year return. The first quintile represents the smallest market capitalization group (small stocks) and the fifth quintile represents the largest market capitalization group (large stocks) with respect to size. Similarly, the first quintile represents the lowest book-to-market ratio (growth stocks) and the fifth quintile represents the highest book-to-market ratio (value stocks) with respect to book-to-market ratio. Finally, the first quintile represents the lowest return (contrarian stocks) and the fifth quintile represents the highest return (momentum stocks) with respect to their past one-year return. When determining the quintiles, only stocks listed on the TSE First Section are included, while stocks with no or negative actual equity capital and those with no past one-year return are excluded. Additionally, stocks newly listed after September are excluded until the next sort month, which is the last business day of August of the following year.

Following these procedures, each stock is assigned a five-level style index ranging from one to five for size, book-to-market ratio, and past one-year return, respectively. Furthermore, based on each fund's holding data, the monthly style index for each fund is calculated by weighting each stock's style index according to its holding weight. Appendix Figure B shows the distribution of the style indices and the number of holdings for each fund, averaged across the time-series.

Appendix Figure B

Distribution of Each Fund's Style Indices and Number of Holdings



Note: Distribution of each fund's monthly style indices and monthly number of holdings, averaged over the investment period.

Appendix C

Hypothetical and Actual Returns

Since the gross returns of the funds, obtained by summing the CS, CT, and AS measures, are hypothetical returns calculated using stock holding data and differ from the actual returns, we examined the deviation between the hypothetical and actual returns.

The respective returns are shown in Appendix Table C. On average, for all funds, the actual returns are about 1% per year lower than the hypothetical returns. This difference is expected since the hypothetical returns are calculated before the deduction of trust fees, whereas the actual returns are after such deductions. However, this difference is smaller than the trust fees. Factors contributing to this discrepancy, aside from trust fees, include the exclusion of transaction fees and other costs from hypothetical returns, the infrequency of updates to stock holding data, and the fact that hypothetical returns do not fully capture actual investment performance due to the presence of stocks for which performance measures cannot be calculated and cash holdings.

In each category, the differences generally align with trust fees, but for mid-small growth funds, the actual returns are higher, suggesting that actual returns may be significantly inflated by transactions not captured by the stock holding data. Given the impact of this factor on the average for all funds, we judged that the hypothetical returns generally reflect the actual management of the funds.

Appendix Table C

Hypothetical and Actual Returns, January 2009–December 2021

Category	Hypothetical return (gross, %/year)	Actual return (net, %/year)	Difference	Trust fee (%/year)
Large blend funds	10.238	8.851	1.387	1.520
Large value funds	10.580	8.541	2.038	1.396
Large growth funds	11.530	10.578	0.952	1.649
Mid-small blend funds	11.792	10.417	1.375	1.578
Mid-small value funds	11.391	10.035	1.356	1.647
Mid-small growth funds	13.687	13.948	-0.261	1.702
Specific region/sector funds	10.806	9.276	1.530	1.427
All funds	11.397	10.403	0.995	1.574

Note: Hypothetical returns are the sum of CS, CT, and AS measures. Actual returns are the rate of return for reinvested dividends before tax and after deduction of trust fees. The analysis period for large value funds is from February 2009 to December 2021, as there were no applicable funds in January 2009.

Appendix D

Factor Returns for FFC4

The factor returns used to estimate α_FFC4 in this paper's analysis are based on the Fama-French 3-Factor Model and the Momentum Factor, which are derived from a database related to Japanese listed stocks provided by FDS.[16]

For the method of creating the factor returns, please refer to the data specifications in footnote 16. To be consistent with the benchmark construction method based on the approach of Daniel *et al.* (1997), the sort universe was defined as the TSE First Section (including financials). Additionally, the momentum factor was calculated using the returns over the past 12 months, ending two months prior.

Appendix E

Additional Analysis Results Using Stock Holding Data

For the benchmarks used to calculate the CS, CT, and AS measures, the quintile points are determined by restricting the selection to stocks listed on the TSE First Section, as described in Appendix A. Although this approach is consistent with the method used to create factor returns when estimating α_FFC4, we recognize the following issues with the method of dividing by market capitalization.

First, as shown in the distribution by size in Appendix Figure B, the mutual funds analyzed are notably skewed toward large-cap stocks, indicating that the large-cap stocks in which mutual funds typically invest are not evenly distributed across the various quintiles.

[16] https://fdsol-services.com/academic/acproduct-list/ (in Japanese)

Additionally, given the long-tailed distribution of market capitalization in the Japanese stock market, the larger the market capitalization quintile, the greater the variation in market capitalization among the stocks within each quintile.

To mitigate these effects, the market capitalization quintiles were determined not only by considering stocks listed on the TSE First Section, but also by imposing an additional condition that included only stocks with a market capitalization of 50 billion yen or more, which is generally regarded as the investment universe for mutual funds.[17] The results, presented in Appendix Table E, indicate that the trends for each measure do not change significantly.

Appendix Table E

Performance Measures, January 2009–December 2021

Category	Number of funds	CS (%/year)	CT (%/year)	AS (%/year)
Large blend funds	125	0.878*	0.326	10.093
		(1.955)	(0.880)	
Large value funds	27	0.949**	0.515	10.191
		(1.998)	(1.251)	
Large growth funds	61	1.353**	0.259	10.837
		(2.467)	(0.628)	
Mid-small blend funds	44	1.433***	0.307	10.904
		(3.033)	(1.162)	
Mid-small value funds	11	1.560***	0.019	10.803
		(3.367)	(0.068)	
Mid-small growth funds	84	2.665***	0.038	11.881
		(3.157)	(0.124)	
Specific region/sector funds	34	1.121*	-0.095	10.798
		(1.893)	(-0.284)	
All funds	385	1.428***	0.191	10.707
		(3.087)	(0.667)	

Note: CS is estimated using equation (2), CT using equation (3), and AS using equation (4). All values are annualized. The analysis period for large value funds is from February 2009 to December 2021, as there were no applicable funds in January 2009. Figures in parentheses represent *t*-values, with * indicating statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Appendix F

Persistence of Performance Measures

The respective performance measures for the decile portfolios sorted by CS, CT, and AS measures over the past one, three, and five years, both at the time of sorting and thereafter, are as follows. The CS measure shows a certain degree of persistence.

[17] In short, stocks listed on the TSE First Section and with a market capitalization of 50 billion yen or more were equally allocated to each quintile.

Appendix Table F-1

CS Measures for Decile Portfolios, January 2013–December 2021

	Sort by past 1-year CS		Sort by past 3-year CS		Sort by past 5-year CS	
Decile portfolios	At sorting (%/year)	Post-sorting (%/year)	At sorting (%/year)	Post-sorting (%/year)	At sorting (%/year)	Post-sorting (%/year)
1 (low)	-6.926	0.848	-3.814	1.364	-2.711	1.177
		(0.798)		(1.433)		(1.244)
1/2	-4.882	0.967	-2.589	1.310*	-1.776	1.259
		(1.094)		(1.717)		(1.584)
3/4	-0.763	1.395**	-0.229	1.291**	0.003	1.346**
		(2.301)		(2.193)		(2.338)
5/6	1.117	1.066**	0.781	1.117**	0.796	1.382**
		(2.096)		(2.084)		(2.460)
7/8	3.087	1.395***	1.921	1.279**	1.786	1.625***
		(2.709)		(2.404)		(3.113)
9/10	8.376	2.660***	5.193	2.533***	4.633	2.566***
		(3.302)		(3.297)		(3.359)
10 (high)	11.267	3.584***	6.969	3.134***	6.101	3.117***
		(3.394)		(2.964)		(3.040)
10-1 spread	18.193	2.736**	10.782	1.770	8.812	1.940*
		(2.029)		(1.582)		(1.744)

Note: Figures in parentheses represent *t*-values, with * indicating statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Appendix Table F-2

CT Measures for Decile Portfolios, January 2013–December 2021

	Sort by past 1-year CT		Sort by past 3-year CT		Sort by past 5-year CT	
Decile portfolios	At sorting (%/year)	Post-sorting (%/year)	At sorting (%/year)	Post-sorting (%/year)	At sorting (%/year)	Post-sorting (%/year)
1 (low)	-4.052	0.279	-3.814	0.176	-1.882	0.288
		(0.661)		(0.394)		(0.671)
1/2	-3.032	0.148	-2.589	-0.016	-1.454	0.041
		(0.410)		(-0.044)		(0.116)
3/4	-0.961	0.061	-0.229	-0.060	-0.510	-0.195
		(0.163)		(-0.175)		(-0.637)
5/6	0.019	-0.252	0.781	-0.089	-0.030	-0.418
		(-0.676)		(-0.227)		(-1.329)
7/8	0.988	-0.272	1.921	-0.362	0.395	-0.280
		(-0.683)		(-0.887)		(-0.884)
9/10	3.025	-0.399	5.193	-0.210	1.247	-0.430
		(-0.865)		(-0.491)		(-1.312)
10 (high)	4.026	-0.333	6.969	-0.158	1.647	-0.430
		(-0.627)		(-0.311)		(-1.106)
10-1 spread	8.078	-0.612	10.782	-0.334	3.529	-0.718
		(-0.940)		(-0.578)		(-1.422)

Note: Figures in parentheses represent *t*-values, with * indicating statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Appendix Table F-3
AS Measures for Decile Portfolios, January 2013–December 2021

	Sort by past 1-year AS		Sort by past 3-year AS		Sort by past 5-year AS	
Decile portfolios	At sorting (%/year)	Post-sorting (%/year)	At sorting (%/year)	Post-sorting (%/year)	At sorting (%/year)	Post-sorting (%/year)
1 (low)	6.049	11.712	8.551	14.236	10.587	9.435
1/2	7.066	11.425	9.026	13.968	10.910	9.277
3/4	9.447	10.806	10.177	14.037	11.730	9.114
5/6	10.894	10.671	10.968	13.919	12.353	9.372
7/8	12.527	10.359	11.951	13.748	13.130	9.475
9/10	16.065	11.110	14.407	14.182	15.123	9.740
10 (high)	17.620	11.011	15.514	14.706	16.077	9.977
10-1 spread	11.571	-0.701	6.963	0.470	5.490	0.542
		(-0.336)		(0.204)		(0.193)

Note: Figures in parentheses represent t-values, with * indicating statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.