11-2001

Persistence in style-adjusted mutual fund returns

Melvyn TEO
Singapore Management University, melvynteo@smu.edu.sg

Sung-Jun WOO
Harvard University

Persistence in Style-Adjusted Mutual Fund Returns

Melvyn Teo[#] and Sung-Jun Woo[*]

This Draft: November 2001

Abstract

The literature on mutual fund persistence took a hit with the finding that one-year stock momentum and expense ratios account for most of the persistence in mutual fund performance (Carhart, 1992; Carhart, 1997). However, since equity mutual funds are grouped into styles (e.g., large value, small growth, mid-cap growth, etc.) and are often confined to trading stocks within their style, one should measure fund performance relative to style when investigating managerial ability. Using CRSP mutual fund data and a methodology similar to Carhart (1997), we find that differences in style-adjusted fund returns persist for up to six years. Neither one-year momentum nor expense ratios explain our results. Our results are also robust to controlling for size, book-to-market equity, load, and total net assets. Since manager tenure is about four years, our results suggest that managerial ability may not be as dead as it seems.

[#*] Department of Economics, Harvard University. [#] E-mail: steo@kuznets.fas.harvard.edu. [*] E-mail: swoo@kuznets.fas.harvard.edu. We thank Kenneth Froot for many helpful comments and suggestions.

1 Introduction

Academic interest in mutual fund return persistence has cooled with the finding that short-term mutual fund persistence is largely driven by fund managers accidentally holding past winner/loser stocks (Carhart, 1997) and the finding that long-term mutual fund persistence is largely driven by persistence in expense ratios (Carhart, 1992). These findings by Carhart suggest that mutual fund persistence is driven neither by differences in mutual fund manager ability nor by mutual fund manager information. They have also put a damper on post-1997 mutual fund return persistence research.

Prior to the influential work by Carhart, research on mutual fund return persistence has been thriving under the expectation that much of it is manager driven. Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), and Wermers (1996) show that mutual fund performance persists in the short term (one to three years). They credit their results to the existence of "hot hands" or common investment strategies. Meanwhile, Grinblatt and Titman (1992), Elton, Gruber, Das, and Hlavka (1993), and Elton, Gruber, Das, and Blake (1996) find that mutual fund performance is predictable over long horizons of between five and ten years. They ascribe their results to mutual fund managers having different information or different stock-picking ability.

By using a 4-factor model which adds to the Fama and French (1993) 3-factor model a momentum factor (PR1YR) that captures the Jegadeesh and Titman (1993) one-year momentum anomaly, Carhart (1997) shows that much of the persistence in one-year mutual fund return can be explained by differential factor loadings on PR1YR. The top decile funds sorted by past one-year returns load positively on PR1YR while bottom decile funds load negatively on PR1YR. Moreover, the 4-factor model explains more than half of the spread in mean excess returns between the funds in these deciles. Further, he shows that the funds do not maintain their decile rankings and as a result, the spread in mean excess returns is not due to the top decile funds following momentum strategies. Rather it is due to their accidentally holding past year winners. Finally, he accounts for most of the remaining spread with differences in expense ratios and transactions costs between the top and bottom deciles sorted on past one-year returns.

However, since equity mutual funds are increasingly identifying themselves by their Morningstar styles[1] (e.g. small value, large growth, etc.) and are often confined to trading stocks within their style, style-adjusted fund returns would be a more appropriate measure of fund manager performance. In his book *Common Sense on Mutual Funds*, John C. Bogle of Vanguard Funds notes that "*style purity* has become the catchphrase of portfolio managers, investment advisers, and mutual fund investors... [M]anagers of individual stock funds today feel pressured to... confine themselves to a given portfolio style that defines the funds' strategy–growth stocks versus value stocks, for example, or large-cap stocks versus small-cap stocks." Clearly, a mutual fund's performance needs to be measured relative to the performance of the average fund in its style.[2] Some authors have adopted a characteristics-based approach to try to benchmark managers' performance relative to their style (Wermers, 2000). This approach assumes that the variation in style returns is entirely induced by variation in the characteristics of the stocks belonging to the style. Teo and Woo (2001) show that past style returns have predictive power over future stock returns in excess of the predictive power of individual stock characteristics like size and book-to-market. This suggests that the variation in style returns is not spanned by variation in stock characteristics across styles. Hence, a style benchmark rather than a stock characteristics benchmark would be more appropriate for investigating mutual fund performance.[3]

Further, one-year fund performance is probably a noisy signal of fund managerial ability. Managers who end up at the top decile of funds sorted on their past one-year returns may be riding on a lucky streak last year. Longer formation periods are needed to pick out the managers (if any) who consistently outperform or underperform the pack. This point was not lost to Carhart. He investigates the returns of mutual funds sorted on their past returns over the prior one-, two-, three-, four-, and five-year periods. He finds scant evidence of mutual fund return persistence in excess of the 4-factor model, expenses, and transactions costs. However, a stricter test of long-term mutual fund performance would be to sort mutual funds based on their returns over a sufficiently long

[1] For instance, T. Rowe Price Small Cap Value fund.

[2] We further discuss the need to measure performance relative to style in Section 3.

[3] Brown and Goetzmann (1995) briefly examine the persistence properties of Wiesenberger style-adjusted fund returns. However, their sample suffers from survivorship bias and they use neither the 3-factor model nor the 4-factor model to test performance persistence. Moreover, the Morningstar styles sort funds more finely than do Wiesenberger codes (Maximum capital gains, Growth, Income, and Growth and Income). Hence a fund's Wiesenberger code is likely to convey less information than its Morningstar style.

period (e.g., three years) to reduce the noise and measurement error due to taking shorter-period performance, and investigate their returns *after* a lag so as to separate away from any short-term persistence stories. For example, we could investigate the one-year returns of mutual funds sorted on the sum of their past two- to four-year returns, three- to five-year returns, and four- to six-year returns.[4]

In this paper, we investigate the properties of style-adjusted mutual fund returns. Style-adjusted mutual fund returns are fund returns in excess of the returns of the average fund in their style. First, we sort funds based on the sum of their past one- to three-year style-adjusted returns to investigate short-term mutual fund performance persistence. Next, since manager tenure is about four years (Chevalier and Ellison, 2000) and if managerial ability is responsible for the persistence in mutual fund performance, style-adjusted performance differences should persist for up to four years. Hence we sort funds based on the sum of their past two- to four-year, three- to five-year, four- to six-year, and five- to seven-year style-adjusted returns to investigate long-term mutual fund performance persistence. We find statistically significant spreads in 4-factor style-adjusted alphas of about 4% per year between funds in the top and bottom deciles when sorted on past one- to three- year performance. This spread decays very slowly when we increase the lag between the formation period and the evaluation period.[5] When funds are sorted on the sum of their past four- to six-year style-adjusted returns the spread is still a healthy 3%. In contrast, the corresponding spreads for sorts on non-style-adjusted returns exhibit a higher rate of decay and higher variance. When funds are sorted on the sum of their past four- to six-year non-style-adjusted returns, the spread is insignificantly different from zero. This may explain why evidence for long-term mutual fund persistence is lacking in the literature. Expense ratios and transactions costs do little to explain the results. The results are suggestive of managerial stock-picking ability.

The rest of the paper is organized as follows. Section 2 describes the data used. Section 3 shows why one should use the style-adjusted metric when investigating fund managerial ability. The empirical results are presented in Section 4 where subsections are devoted to examining 4-factor

[4] The sum of past four- to six-year returns refers to the sum of past annual returns four calendar years ago, five calendar years ago, and six calendar years ago.

[5] Funds are sorted into portfolios based on their performance in the formation period. The performance of these portfolios are then tested in the evaluation period.

and cost-based explanations of mutual fund performance persistence. The consistency properties of style-adjusted return fund rankings are also tested. Section 5 concludes.

2 Data

We use Center for Research in Security Prices (CRSP) mutual fund data. We obtain our mutual fund data from CRSP rather than from Morningstar, as it is well known that the latter has survivorship bias. Morningstar data only include surviving funds. As a result, Elton, Gruber, and Blake (1996) have shown that this causes overall performance measures to be inflated between 40 basis points and 1% per year, depending on the length of the sample period. Other researchers who use CRSP mutual fund data include Carhart (1997), Zheng (1999), and Wermers (2000).[6]

For styles, we use Morningstar equity styles which categorize equity funds into small, midcap, and large, and value, blend, and growth simultaneously. Morningstar styles are used as funds have increasingly begun to identify themselves by their Morningstar equity styles. Some funds that do so include IAI Midcap Growth Fund, BlackRock Large Cap Value Equity Fund, and Nicholas-Applegate Small Cap Growth Fund. Numerous other examples abound. Teo and Woo (2001) use Morningstar equity style return and flow information to show that style information have explanatory power on future stock returns over and above that of a stock's size and book-to-market. This is true even though size and book-to-market equity are essentially determinants of a stock's inclusion in a style. The point to note is that style information is highly discrete while stock characteristics are much more continuous. Two stocks with similar book-to-market equity ratios may be subject to very different forces by virtue of one (with the higher book-to-market equity) being held more by value funds and the other (with a lower book-to-market equity) being held more by blend funds.

Since CRSP mutual fund data do not have Morningstar equity style information, we transfer historical Morningstar style data from the Morningstar Prinicipia Pro Plus CD (Feb 2001) and from the Morningstar Mutual Fund manuals (1993 to 1999) by hand onto the CRSP database.[7]

[6] According to Elton, Gruber, and Blake (2001), returns in the CRSP database for months with multiple distributions on the same day are overstated. This problem appears to have been corrected by CRSP.

[7] Following the standard practice in the mutual fund literature, we omit international funds, sector funds, and domestic hybrid funds. Also, there are concerns that for such funds the Fama-French factors may not adequately

As Morningstar style information is not available prior to 1993, we assume that funds that have existed prior to 1993 operated under their 1993 style in the years before 1993. According to Warther (1984), "mutual funds played a much smaller role in the pre-1984 markets." Hence, our aim is to see if style-adjusted mutual fund returns beginning January 1984 can be explained with past style-adjusted mutual fund returns. To do so, we use mutual fund data beginning January 1977 so as to accommodate a seven-year lag.

Note that not all mutual funds in the CRSP database are featured in the Morningstar databases from 1993 to 1999 and style information is not available for funds that died before 1993. Of the 32,492 fund-years that we have in our database, we do not have style information on 5,364 fund-years or about 16.5% of the fund-years. One way to deal with this is to throw out these fund-years. But that would introduce a survivorship bias into the mutual fund data since many of these fund-years belong to non-surviving pre-1993 funds. Instead, we construct an algorithm based on the informativeness of the fund names, the Wiesenberger fund type code, the ICDI fund objective code, and Strategic Insight's fund objective code reported in the CRSP mutual fund database to estimate the funds' styles. Details of the algorithm are available in the Appendix. Basically, the algorithm takes advantage of simple facts. For instance, funds classified by Strategic Insight as income growth funds (which are often heavily invested in large, dividend-paying stocks) are usually large value funds, and a fund named Munder Mid-Cap Growth Fund would most likely be a mid-cap growth fund.

3 Why Style-Adjusted?

Mutual fund performance studies such as Carhart (1997) have traditionally sorted mutual funds on their past returns. This method overlooks a simple fact. The returns of different mutual fund styles vary considerably. By sorting on mutual fund returns, the researcher not only extracts funds with above-average managers (if any) but also extracts funds in above-average performing styles. For instance, funds sorted into the best return decile are both good funds and not-so-good funds which by chance belong to a well-performing style last year. If the researcher's aim is to make inferences on fund managerial ability, then her sample of above-average return funds is contaminated. In other

cover the risks involved.

Figure 1

Probability of a fund being in style i conditional on being in decile j (where a fund's inclusion in decile j is determined by its current-year fund returns)

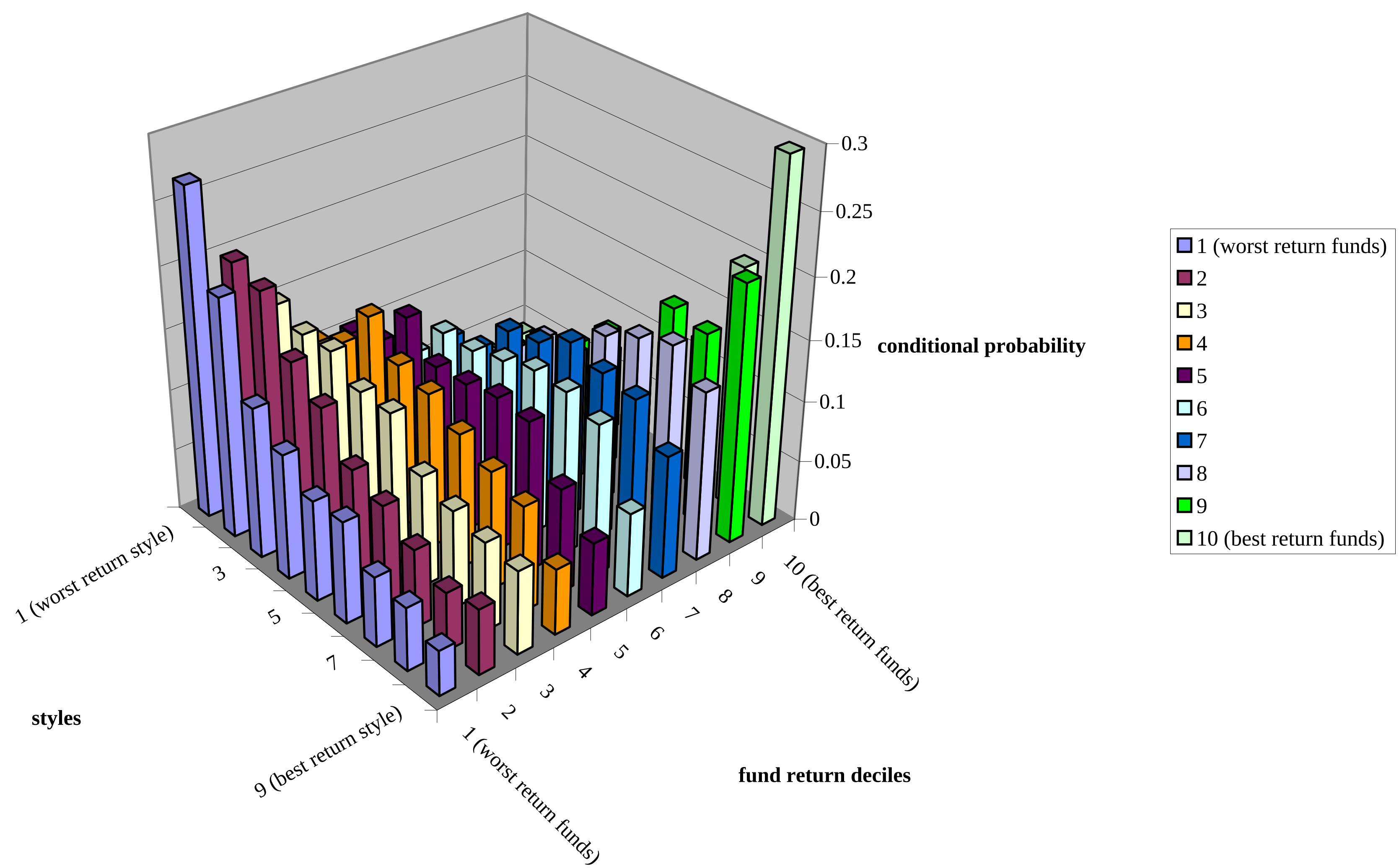
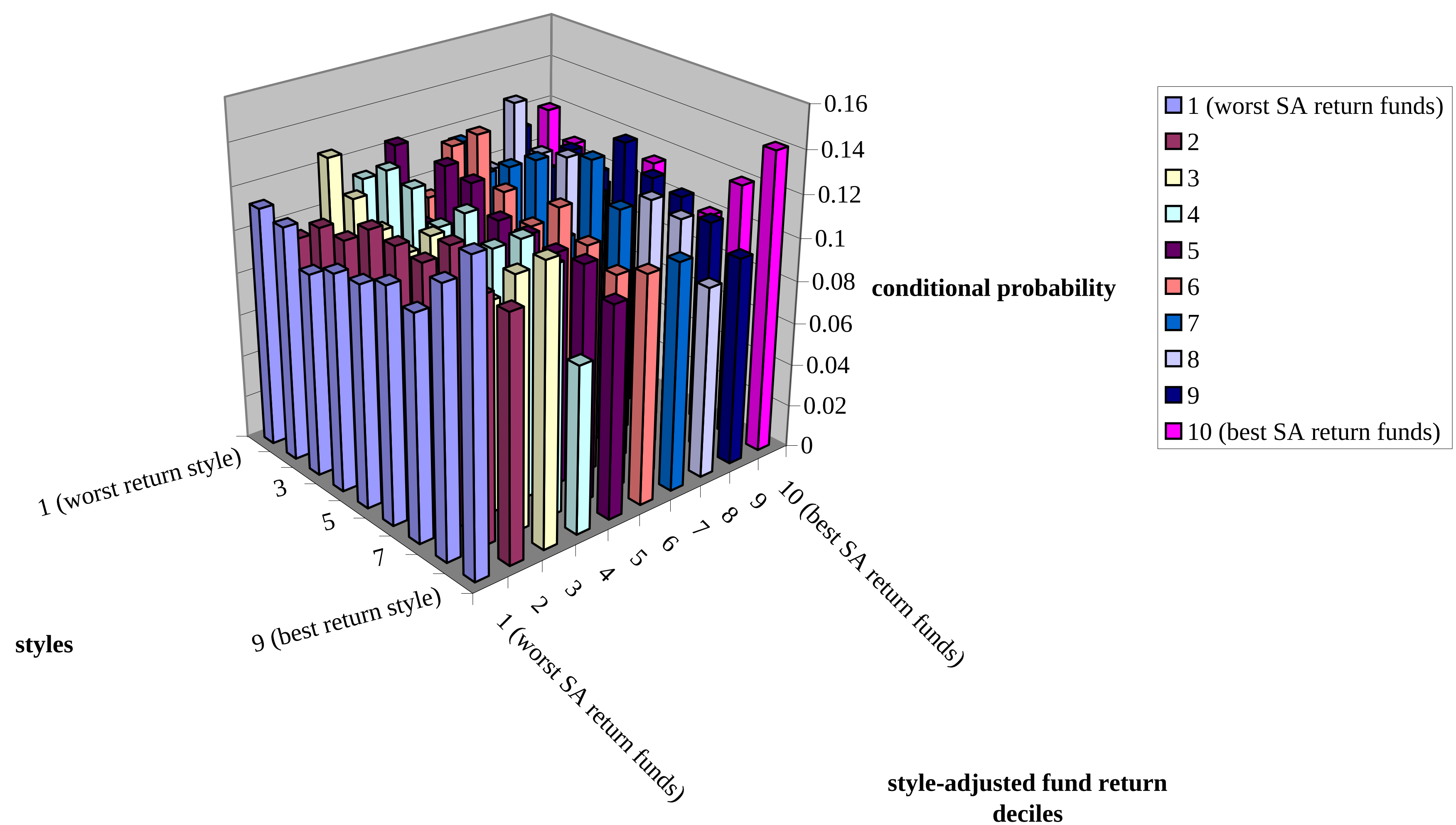


Figure 2
Probability of a fund being in style i conditional on being in decile j (where a fund's inclusion in decile j is determined by its current-year style-adjusted (SA) returns)



words, fund performance includes two components: managerial ability and style performance, and by looking at fund return persistence, the researcher fails to control for the latter, which is unlikely the result of a manager's actions since funds are increasingly identifying themselves by their styles.

Consistent with this intuition, we find that funds in the worst return deciles belong overwhelmingly to the worst performing styles. Conversely, funds in the best return deciles belong overwhelmingly to the best performing styles. Figure 1 graphs the probability of a fund being in a style conditional on being in a certain return decile where its inclusion in a decile is determined by its current-year returns. It shows that funds in the best and worst deciles have a greater-than-25% probability of belonging to the best and worst styles respectively. Moreover, for the worst return decile, the conditional probabilities decrease almost monotonically from the worst return style to the best return style. The analogous phenomenon occurs with the best return decile: conditional probabilities increase almost monotonically from the worst return style to the best return style. Further, there is an almost clinical correlation between the return rank of a fund decile and the return rank of the style to which the funds in the decile most likely belong. This suggests that the return of the style to which a fund belongs is an important determinant of a fund's inclusion into a return decile. Since style return rankings vary considerably from year to year, by using the return metric, a researcher would infer that one year most of the good managers are in small value and another year most of the good managers are in large growth.[8] This does not jibe well with the a priori assumption that managerial ability should be roughly uniform across styles.[9]

When funds are sorted on style-adjusted returns, they do not exhibit a tendency to cluster into certain well-performing or poorly-performing styles. Figure 2 depicts the probability of a fund being in a style conditional on being in a certain decile where its inclusion in a decile is determined by its current-year style-adjusted returns. It reveals a lack of pattern in conditional probabilities which starkly contrasts with the anomalous pattern of conditional probabilities observed in Figure 1. Hence when investigating managerial ability, style-adjusted returns should be the performance

[8] For example, in 1995, most of the funds in the loser return decile are small value funds, yet most of the funds in the same decile next year are mid-cap growth funds.

[9] In addition, for each year, we perform Spearman rank tests on the return rank of a style and the percentage of funds, from that style, that make up the best fund return decile. The Spearman rank tests reject the hypothesis that the style return rank and the percentage of funds are independent at the 10% level for 13 out of the 16 years in our sample. The same result prevails when we consider funds in the worst return decile.

metric on which the researcher bases her inferences. This is true whether or not fund returns or style-adjusted fund returns yield greater evidence for persistence.[10]

4 Empirical Results

4.1 Style-Adjusted Returns and Common Factors in Stock Returns

In this section, we investigate the properties of style-adjusted fund returns by taking into account common factors in stock returns. We form portfolios of funds based on three-year sums of their past style-adjusted returns. A three-year formation period is used as a sort on one-year fund performance data may not be adequate in separating luck from potential managerial ability. We then estimate the performance of the resulting portfolios after various lags and follow the methodology of Carhart (1997). On January 1 of each year, we form ten equal-weighted portfolios of funds, using different sums of past three-year style-adjusted fund returns, e.g., the sum of past one- to three-year past returns, the sum of two- to four-year past returns, etc. We hold the portfolios for one year, then re-form them. This yields a time series of monthly returns on each portfolio from January 1984 to December 1999. Funds that disappear during the course of the year are included in the equal-weighted average until they disappear, then the portfolio weights are readjusted appropriately.

We employ three models of performance measurement: the Capital Asset Pricing Model (CAPM) described in Sharpe (1964) and Lintner (1965), the Fama and French (1993) 3-factor model, and the Carhart (1997) 4-factor model. Carhart's 4-factor model adds to Fama and French's 3-factor model a momentum factor that captures Jegadeesh and Titman's (1993) one-year momentum anomaly. We shall not discuss these models in detail but instead direct the reader to Carhart (1997) who gives an excellent review. The advantages of the 3-factor model relative to the CAPM in explaining the cross-sectional variation of stock returns are well-known and we will not belabor the point. Also,

[10] One predictable critique is that since funds are sorted into Morningstar styles based on the size and book-to-market (and price-to-earnings) ratio of their holdings, funds sorted on Fama-French 3-factor alpha will not crowd into well-performing or poorly-perfoming styles in a systematic fashion. This is not true. When the alpha analogue of Figure 1 is plotted, we find that funds in the best alpha decile crowd into styles with the best alpha. Conversely, funds in the worst alpha decile crowd into styles with the worst alpha. The alpha analogue of Figure 1 easily passes the eyeball test as far as displaying a positive relationship between style alpha ranking and fund alpha ranking is concerned.

Table 1

Portfolios of Mutual Funds Formed on Three-Year Sums of Past Style-Adjusted Fund Returns

Mutual funds are sorted on January 1 each year from January 1984 to December 1999 into ten portfolios based on the three-year sum of their returns minus their styles' returns. The portfolios are equally weighted monthly so the weights are re-adjusted whenever a stock disappears. Spread A is the return of the funds in decile 1 minus the return of the funds in decile 10. Spread B is the return of the funds in decile 2 minus the return of the funds in decile 10. Spread C is the return of the funds in decile 1 minus the returns of the funds in decile 9. 123spreadA denotes the spread A derived from a sort on sum of past one-, two-, and three-year style-adjusted returns. Similarly 234spreadA denotes the spread A derived from a sort on the sum of past two-, three-, and four-year style-adjusted returns. VWRF is the excess return on the CRSP value-weighted market proxy. RMRF, SMB, and HML are Fama and French's (1993) market proxy and factor-mimicking portfolios for size and book-to-market equity. PR1YR is Carhart's (1997) factor-mimicking portfolio for one-year return momentum. Alpha is the intercept of the Model. The t-statistics are in parentheses. The number of observations for each regression is 192.

	Monthly		CAPM			Fama-French 3-Factor Model					Carhart 4-Factor Model					
	Excess				Adj					Adj						Adj
Portfolio	Return	Std Dev	Alpha	VWRF	R-sq	Alpha	RMRF	SMB	HML	R-sq	Alpha	RMRF	SMB	HML	PR1YR	R-sq
123spreadA	0.39%	1.14%	**0.37**	0.02	-0.002	**0.40**	-0.01	-0.08	-0.12	0.060	**0.33**	-0.01	-0.05	-0.11	0.06	0.084
			(4.42)**	(0.83)		**(4.77)****	(-0.68)	(-2.43)	(-3.46)		**(3.85)****	(-0.63)	(-1.54)	(-3.03)	(2.41)	
123spreadB	0.36%	0.95%	**0.36**	0.00	-0.005	**0.35**	-0.01	-0.11	-0.07	0.074	**0.30**	-0.01	-0.09	-0.06	0.05	0.095
			(5.12)**	(0.02)		**(5.14)****	(-0.58)	(-4.06)	(-2.37)		**(4.25)****	(-0.53)	(-3.14)	(-1.96)	(2.28)	
123spreadC	0.19%	0.87%	**0.18**	0.01	-0.000	**0.22**	-0.02	-0.02	-0.13	0.094	**0.19**	-0.02	-0.01	-0.12	0.03	0.099
			(2.77)**	(0.97)		**(3.53)****	(-1.41)	(-0.93)	(-4.67)		**(2.93)****	(-1.38)	(-0.42)	(-4.35)	(1.43)	
234spreadA	0.28%	1.01%	**0.27**	0.01	-0.004	**0.29**	-0.02	-0.07	-0.11	0.062	**0.28**	-0.02	-0.07	-0.11	0.01	0.057
			(3.59)**	(0.55)		**(3.92)****	(-0.94)	(-2.55)	(-3.52)		**(3.65)****	(-0.94)	(-2.32)	(-3.41)	(0.24)	
234spreadB	0.32%	0.92%	**0.33**	-0.01	-0.001	**0.33**	-0.02	-0.11	-0.07	0.087	**0.32**	-0.02	-0.10	-0.07	0.01	0.082
			(4.93)**	(-0.93)		**(4.95)****	(-1.41)	(-4.29)	(-2.39)		**(4.58)****	(-1.40)	(-3.92)	(-2.28)	(0.38)	
234spreadC	0.09%	0.74%	0.05	0.05	0.091	0.08	0.02	-0.01	-0.10	0.168	0.09	-0.02	-0.10	-0.07	0.01	0.167
			(0.87)	(4.48)		(1.57)	(1.83)	(-0.53)	(-4.40)		(1.72)	(-1.40)	(-3.92)	(-2.28)	(0.38)	
345spreadA	0.23%	0.87%	**0.19**	0.05	0.049	**0.22**	0.02	-0.03	-0.10	0.112	**0.21**	0.02	-0.03	-0.10	0.00	0.107
			(2.99)**	(3.31)		**(3.53)****	(1.14)	(-1.37)	(-3.90)		**(3.31)****	(1.14)	(-1.25)	(-3.80)	(0.13)	
345spreadB	0.25%	0.74%	**0.24**	0.00	-0.005	**0.25**	-0.01	-0.07	-0.07	0.060	**0.25**	-0.01	-0.07	-0.07	0.00	0.055
			(4.54)**	(0.04)		**(4.72)****	(-1.01)	(-3.21)	(-2.98)		**(4.52)****	(-1.01)	(-3.07)	(-2.94)	(-0.11)	
345spreadC	0.04%	0.68%	0.00	0.05	0.0815	0.03	0.02	0.02	-0.07	0.149	0.06	0.02	0.01	-0.08	-0.02	0.156
			(-0.06)	(4.24)		(0.67)	(1.76)	(0.99)	(-3.93)		(1.11)	(1.73)	(0.44)	(-3.85)	(-1.53)	
456spreadA	0.25%	0.94%	**0.19**	0.06	0.068	**0.24**	0.02	-0.03	-0.14	0.166	**0.25**	0.02	-0.04	-0.14	-0.01	0.163
			(2.88)**	(3.85)		**(3.66)****	(1.14)	(-1.30)	(-4.95)		**(3.66)****	(1.12)	(-1.42)	(-4.97)	(-0.57)	
456spreadB	0.20%	0.74%	**0.19**	0.01	-0.004	**0.20**	-0.01	-0.06	-0.06	0.051	**0.22**	-0.01	-0.07	-0.07	-0.02	0.054
			(3.51)**	(0.50)		**(3.60)****	(-0.41)	(-3.10)	(-2.58)		**(3.83)****	(-0.44)	(-3.36)	(-2.77)	(-1.29)	
456spreadC	0.16%	0.81%	0.09	0.07	0.149	**0.13**	0.04	0.04	-0.09	0.236	**0.16**	0.04	0.02	-0.10	-0.02	0.241
			(1.59)	(5.88)		**(2.52)***	(3.04)	(1.75)	(-3.91)		**(2.88)****	(3.01)	(1.16)	(-4.14)	(-1.53)	
567spreadA	0.14%	0.79%	0.07	0.07	0.159	**0.11**	0.04	0.00	-0.10	0.230	**0.12**	0.04	0.00	-0.10	-0.01	0.228
			(1.29)	(6.09)		**(2.05)***	(3.28)	(0.11)	(-4.24)	-2.05	**(2.17)***	(3.25)	(-0.13)	(-4.29)	(-0.72)	
567spreadB	0.11%	0.72%	**0.10**	0.01	-0.003	**0.11**	0.00	-0.03	-0.03	0.000	**0.14**	0.00	-0.04	-0.04	-0.03	0.012
			(1.97)*	(0.62)		**(2.01)***	(0.05)	(-1.28)	(-1.32)		**(2.46)***	(0.01)	(-1.79)	(-1.61)	(-1.77)	
567spreadC	0.06%	0.84%	-0.06	0.13	0.455	-0.02	0.10	0.04	-0.08	0.529	-0.01	0.10	0.04	-0.08	-0.01	0.527
			(-1.36)	(12.65)		(-0.42)	(9.13)	(2.39)	(-4.32)		(-0.27)	(9.10)	(2.12)	(-4.32)	(-0.42)	

**Alpha significant at the 1% level

*Alpha significant at the 5% level

Table 2

Portfolios of Mutual Funds Formed on Three-Year Sums of Past Fund Returns

Mutual funds are sorted on January 1 each year from January 1984 to December 1999 into ten portfolios based on the three-year sum of their returns. The portfolios are equally weighted monthly so the weights are re-adjusted whenever a stock disappears. Spread A is the return of the funds in decile 1 minus the return of the funds in decile 10. Spread B is the return of the funds in decile 2 minus the return of the funds in decile 10. Spread C is the return of the funds in decile 1 minus the returns of the funds in decile 9. 123spreadA denotes the spread A derived from a sort on sum of past one-, two-, and three-year returns. Similarly 234spreadA denotes the spread A derived from a sort on the sum of past two-, three-, and four-year returns. VWRF is the excess return on the CRSP value-weighted market proxy. RMRF, SMB, and HML are Fama and French's (1993) market proxy and factor-mimicking portfolios for size and book-to-market equity. PR1YR is Carhart's (1997) factor-mimicking portfolio for one-year return momentum. Alpha is the intercept of the Model. The t-statistics are in parentheses. The number of observations for each regression is 192.

	Monthly Style-adjusted		CAPM			Fama-French 3-Factor Model					Carhart 4-Factor Model					
Portfolio	Return	Std Dev	Alpha	VWRF	Adj R-sq	Alpha	RMRF	SMB	HML	Adj R-sq	Alpha	RMRF	SMB	HML	PR1YR	Adj R-sq
123spreadA	0.38%	1.79%	**0.38**	0.00	-0.005	**0.42**	-0.05	-0.10	-0.19	0.065	**0.37**	-0.05	-0.08	-0.18	0.05	0.052
			(2.87)**	(-0.15)		(3.22)**	(-1.60)	(-2.02)	(-3.40)		(2.71)**	(-1.57)	(-1.53)	(-3.15)	(1.17)	
123spreadB	0.39%	1.58%	**0.44**	-0.05	0.017	**0.41**	-0.06	-0.24	-0.10	0.156	**0.37**	-0.06	-0.22	-0.09	0.03	0.156
			(3.82)**	(-2.08)		(3.72)**	(-2.19)	(-5.72)	(-2.20)		(3.23)**	(-2.16)	(-5.08)	(-1.99)	(1.00)	
123spreadC	0.13%	1.55%	0.11	0.01	-0.004	0.20	-0.06	0.00	-0.21	0.088	0.17	-0.05	0.02	-0.20	0.03	0.087
			(1.02)	(0.44)		(1.80)	(-1.92)	(0.08)	(-4.43)		(1.46)	(-1.90)	(0.34)	(-4.21)	(0.83)	
234spreadA	0.27%	1.54%	**0.26**	0.01	-0.005	**0.30**	-0.02	-0.04	-0.11	0.012	**0.31**	-0.02	-0.04	-0.12	-0.02	0.008
			(2.31)*	(0.37)		(2.57)*	(-0.73)	(-0.82)	(-2.27)		(2.58)*	(-0.74)	(-0.92)	(-2.31)	(-0.46)	
234spreadB	0.26%	1.34%	**0.30**	-0.05	0.016	**0.26**	-0.04	-0.17	-0.04	0.113	**0.29**	-0.04	-0.18	-0.05	-0.02	0.112
			(3.05)**	(-2.03)		(2.75)**	(-1.67)	(-4.78)	(-1.09)		(2.88)**	(-1.69)	(-4.79)	(-1.22)	(-0.86)	
234spreadC	0.04%	1.42%	-0.01	0.05	0.017	0.05	0.01	0.06	-0.10	0.054	0.10	0.01	0.04	-0.11	-0.05	0.060
			(-0.05)	(2.11)		(0.49)	(0.50)	(1.49)	(-2.16)		(0.93)	(0.46)	(0.93)	(-2.40)	(-1.50)	
345spreadA	0.25%	1.48%	0.21	0.04	0.012	**0.32**	-0.03	0.11	-0.20	0.173	**0.29**	-0.03	0.12	-0.19	0.03	0.173
			(1.95)	(1.80)		(3.22)**	(-1.16)	(2.86)	(-4.63)		(2.76)**	(-1.14)	(3.02)	(-4.39)	(0.98)	
345spreadB	0.17%	1.18%	**0.20**	-0.03	0.004	**0.22**	-0.05	-0.04	-0.09	0.025	**0.22**	-0.05	-0.05	-0.09	-0.01	0.02
			(2.29)*	(-1.36)		(2.49)*	(-2.18)	(-1.34)	(-2.32)		(2.44)*	(-2.18)	(-1.34)	(-2.32)	(-0.23)	
345spreadC	-0.01%	1.41%	-0.09	0.09	0.069	0.03	0.02	0.17	-0.16	0.276	0.04	0.02	0.17	-0.16	-0.01	0.272
			(-0.87)	(3.90)		(0.33)	(0.89)	(4.97)	(-4.12)		(0.43)	(0.87)	(4.56)	(-4.12)	(-0.40)	
456spreadA	0.14%	1.64%	0.08	0.07	0.026	0.20	-0.01	0.14	-0.19	0.170	0.20	-0.01	0.13	-0.19	-0.01	0.166
			(0.64)	(2.45)		(1.74)	(-0.29)	(3.20)	(-3.99)		(1.71)	(-0.30)	(2.96)	(-3.95)	(-0.18)	
456spreadB	0.11%	1.34%	0.12	-0.01	-0.003	0.14	-0.03	-0.02	-0.06	-0.003	0.19	-0.03	-0.03	-0.07	-0.04	0.002
			(1.25)	(-0.60)		(1.41)	(-1.18)	(-0.44)	(-1.41)		(1.78)	(-1.22)	(-0.88)	(-1.64)	(-1.43)	
456spreadC	0.04%	1.60%	-0.06	0.11	0.080	0.10	0.01	0.23	-0.22	0.379	0.11	0.01	0.22	-0.23	-0.01	0.377
			(-0.53)	(4.19)		(1.06)	(0.56)	(6.30)	(-5.46)		(1.13)	(0.55)	(5.80)	(-5.44)	(-0.42)	
567spreadA	-0.06%	1.51%	-0.17	0.12	0.110	-0.09	0.07	0.13	-0.10	0.192	-0.07	0.07	0.11	-0.11	-0.02	0.190
			(-1.65)	(4.95)		(-0.92)	(2.75)	(3.21)	(-2.37)		(-0.67)	(2.73)	(2.81)	(-2.45)	(-0.67)	
567spreadB	-0.06%	1.29%	-0.11	0.06	0.034	-0.11	0.06	0.00	-0.01	0.025	-0.07	0.06	-0.01	-0.02	-0.03	0.025
			(-1.18)	(2.79)		(-1.10)	(2.27)	(0.14)	(-0.21)		(-0.72)	(2.25)	(-0.22)	(-0.39)	(-1.07)	
567spreadC	-0.17%	1.48%	**-0.32**	0.17	0.232	**-0.22**	0.11	0.20	-0.11	0.409	**-0.22**	0.11	0.20	-0.11	0.00	0.406
			(-3.37)**	(7.66)		(-2.54)*	(5.09)	(6.14)	(-2.86)		(-2.44)*	(5.08)	(5.81)	(-2.80)	(0.07)	

**Alpha significant at the 1% level

*Alpha significant at the 5% level

according to Carhart (1997), the 4-factor model "eliminates almost all of the patterns in (3-factor model) pricing errors, indicating that it well describes the cross-sectional variation in average stock returns."

We estimate performance relative to the CAPM, 3-factor, and 4-factor models as

$$SAr_{im} = \alpha_{iM} + \beta_{iM} VWRF_m + e_{im} \quad m = 1, 2, ..., M \tag{1}$$

$$SAr_{im} = \alpha_{iM} + \beta_{iM} VWRF_m + s_{iM} SMB_m + h_{iM} HML_m + e_{im} \quad m = 1, 2, ..., M \tag{2}$$

$$SAr_{im} = \alpha_{iM} + \beta_{iM} VWRF_m + s_{iM} SMB_m + h_{iM} HML_m + p_{iM} PR1YR_m + e_{im} \quad m = 1, 2, ..., M \tag{3}$$

where SAr_{im} is the equal-weighted style-adjusted return of the funds in portfolio i in month m; VWRF is the excess return on the CRSP value-weighted portfolio of all NYSE, AMEX, and Nasdaq stocks; RMRF is the excess return on a value-weighted aggregate market proxy; and SMB, HML, and PR1YR[11] are returns on value-weighted, zero-investment, factor-mimicking portfolios for size, book-to-market equity, and one-year momentum in stock returns, respectively.

The spread in mean monthly style-adjusted returns between the best and worst funds (spreadA)[12] displayed in Table 1 starts at 39 basis points (4.7% per year) for the sort on the sum of one- to three-year style-adjusted returns. It decays slowly as we increase the length of the lag between the formation period and the evaluation period. When funds are sorted on the sum of four- to six-year style-adjusted returns, the mean monthly style-adjusted spread is 25 basis points (3% per year). In contrast, the spread in mean monthly returns displayed in Table 2 (where we perform the corresponding regressions with excess returns[13] in place of style-adjusted returns) decays at a much faster rate as we increase the lag. It starts off at a healthy 38 basis points for the sort on one- to three-year returns and falls to 14 basis points (1.68% per year) for the sort on four- to six-year

[11] Data on PR1YR are generously provided by Mark Carhart. See Carhart (1997) for a detailed description of PR1YR's construction.

[12] Henceforth, unless otherwise mentioned, "spread" refers to the spread between the best and worst funds based on past performance (spreadA).

[13] Excess return is the return minus the one-month T-bill return.

returns. The spread in mean monthly returns also has a higher variance than its style-adjusted counterpart. The standard deviations of the return spreads are uniformly above 140 basis points, while those for the style-adjusted return spreads are uniformly below 115 basis points. This suggests that fund return is a noisier predictor of future fund performance than is style-adjusted fund return.

The CAPM explains neither the spread in style-adjusted returns nor the spread in normal returns. The loadings on VWRF or the betas of the spreads are usually insignificantly different from zero. Since the spread for the sort on style-adjusted returns has a higher precision than that for the sort on returns, the CAPM alphas for the former sort are also overwhelmingly more significant. The CAPM alpha for the style-adjusted spread remains significant at the 5% level up to the four- to six-year sort, while the CAPM alpha for the return spread loses its significance after the two- to four-year sort. In fact, the CAPM alpha for the return spread is *negatively* significant at the 10% level for the five- to seven-year sort. One possible reason for this is that style returns mean revert after four to six years.

The Fama-French 3-factor alphas are also more significant and persistent for the style-adjusted return sort than for the return sort. The 3-factor alpha for the style-adjusted return spread between the best and worst portfolios is positively significant at the 5% level up to and including the five- to seven-year sort. On the other hand, the 3-factor alpha for the return spread is insignificant beyond the three- to five-year sort. The three factors do somewhat explain the spreads, though. The spreads load negatively on HML, implying that the top decile funds hold more growth stocks than do the bottom decile funds. Also, most of the spreads load negatively on SMB. However, since during this period the mean return on HML is positive and that on SMB is negative, and the spreads load more negatively on HML, the 3-factor alphas are generally higher than the CAPM alphas.

Carhart (1997) finds that much of the previously documented one-year persistence in mutual fund returns can be explained by the momentum factor (PR1YR) in his 4-factor model. However, since last year's winner funds frequently become this year's loser funds and vice versa, at least when sorted on returns, he attributes the significant and positive PR1YR loadings to managers in the top decile accidentally holding last year's winner stocks. The 4-factor alphas for sorts on returns displayed in Table 2 broadly concur with this story. The loadings on PR1YR are all insignificantly

Figure 3
Cumulative style-adjusted (SA) returns for portfolios sorted on the sum of past one- to three-year SA fund returns (1-72 months into the evaluation period)

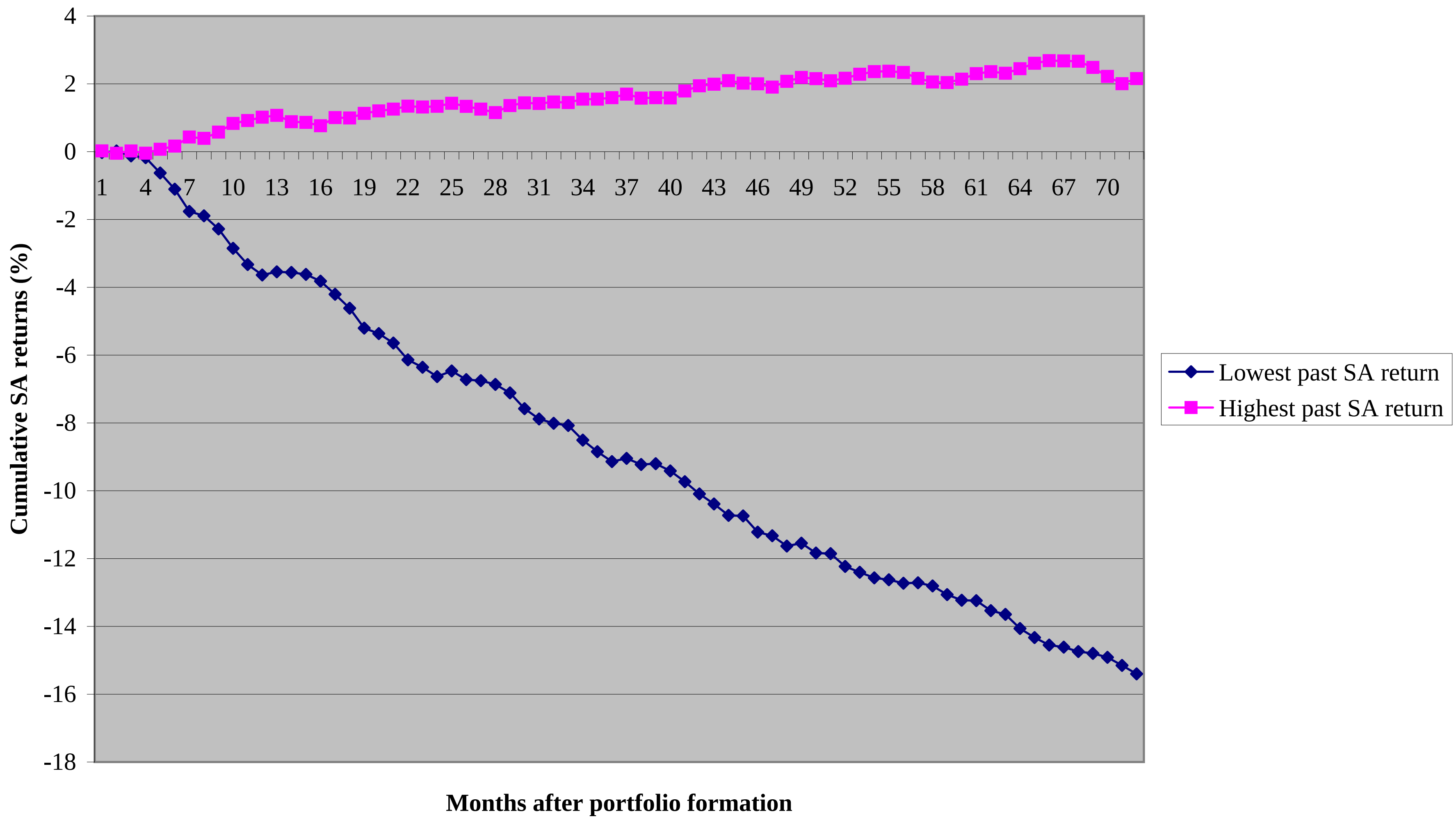
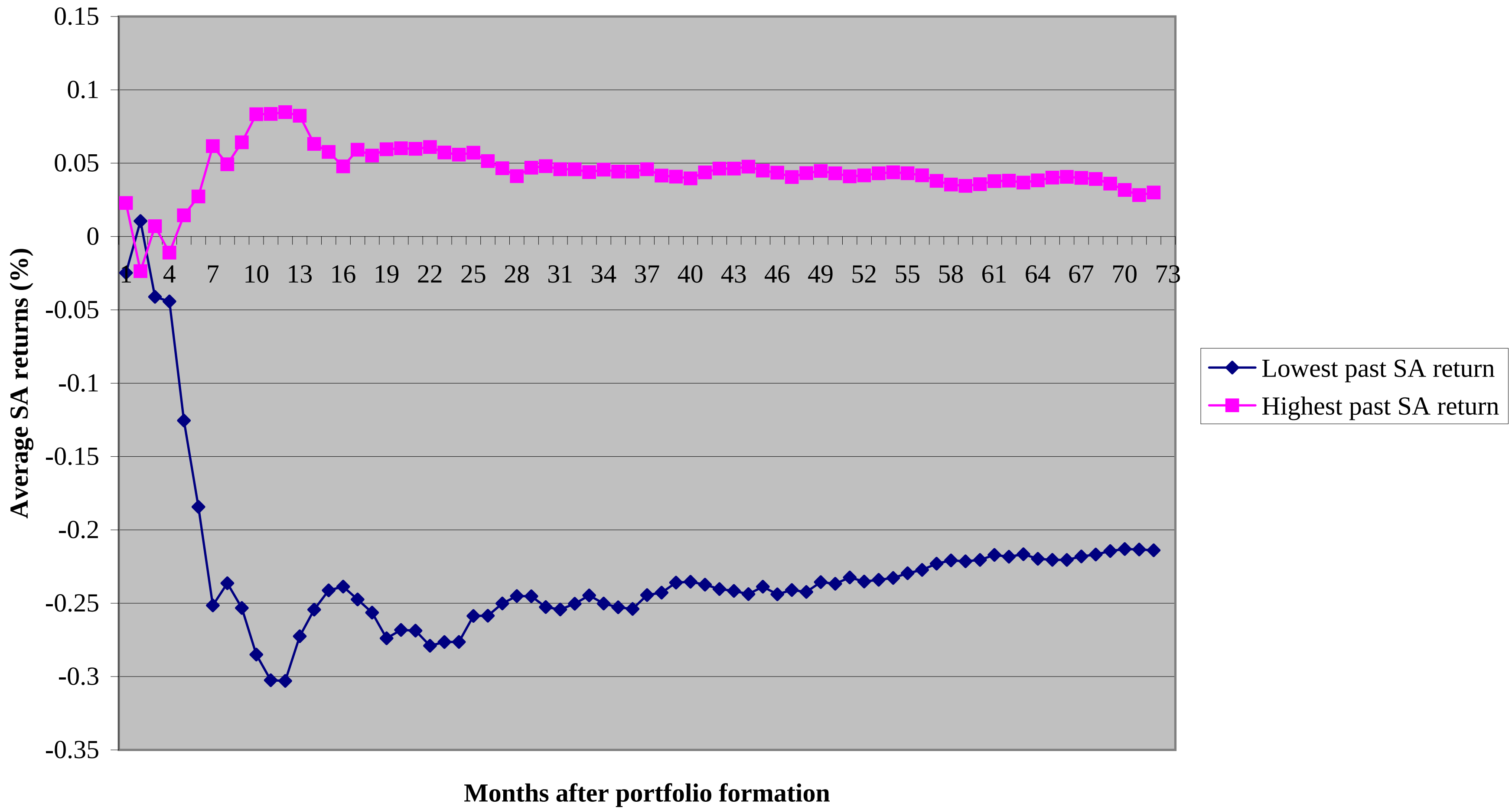


Figure 4
Average monthly holding-period style-adjusted (SA) returns for portfolios sorted on the sum of past one- to three-year SA fund returns (1-72 months into the evaluation period)



different from zero for the two- to four-year sort and beyond. In results not reported, we find that the sort on past year returns generates a spread that loads positively and significantly on PR1YR. Moreover, the spread and all the decile portfolios for the sort on returns two years ago load insignificantly on PR1YR. The same situation prevails for the sort on style-adjusted returns.[14] If managers do adopt momentum strategies consciously, then the positive loadings on PR1YR should persist. Hence the evidence suggests that managers do not consciously adopt momentum strategies.

Because of the predominantly insignificant loadings on PR1YR, the 4-factor alphas for the spreads in Tables 1 and 2 do not differ much from the 3-factor alphas. Again, the 4-factor alphas are more significant for the sorts on style-adjusted returns than for the sorts on returns. Also, they persist more for the former than for the latter. With the sorts on returns, 4-factor alphas become insignificantly different from zero beyond the sort on past three- to five-year returns, while the 4-factor alphas are all significantly positive for the sorts on style-adjusted returns. So far we have compared the spread between deciles 1 and 10 in Tables 1 and 2. The same can be said about the spread between deciles 2 and 10 (spreadB). The spread between deciles 1 and 9 (spreadC), however, is much weaker for both style-adjusted and non-style-adjusted sorts. This reflects the fact that it is the underperformance of the bottom decile that drives much of the results.

Figure 3 depicts the cumulative style-adjusted returns after 1 to 72 months for the top and bottom deciles sorted on the sum of past one- to three-year fund style-adjusted returns. The most striking part of the graph is undoubtedly the asymmetry between the winner and loser portfolios. The underperformance for the loser portfolio is stronger than the overperformance for the winner portfolio. Figure 4, which displays the average monthly holding-period style-adjusted returns for the same sort, suggests that the decay in the spread is slow as the holding period lengthens.

So far, the results in this section have relied on the assumption that the factor loadings for risk are constant over time for each portfolio. However, it is possible that the factor loadings of the portfolios vary through time.[15] Even when we look at the factor loadings of a fund over time,

[14] One may note that the spread loads positively on PR1YR when sorted on the sum of one- to three-year style-adjusted returns. However, this is mostly driven by the negative loading on PR1YR for the bottom decile in the sort on past one-year style-adjusted returns.

[15] We thank Tuomo Vuolteenaho for pointing this out.

they too may vary. For instance, Fama and French (1997) find that the factor loadings of industry portfolios vary through time. To circumvent this problem, we adopt a methodology very similar to that of Fama and French (1997) and estimate the following two conditional regressions each year in place of the standard 3-factor and 4-factor regressions:

$$\begin{aligned} r_{im} &= \alpha_{iM} + \beta_{iM} VWRF_m + [s_{i1M} + s_{i2M} SDUM_{im}] SMB_m \\ &\quad + [h_{i1M} + h_{i2M} VDUM_{im}] HML_m + e_{im} \\ m &= 1, 2, ..., M \end{aligned} \tag{4}$$

$$\begin{aligned} r_{im} &= \alpha_{iM} + \beta_{iM} VWRF_m + [s_{i1M} + s_{i2M} SDUM_{im}] SMB_m \\ &\quad + [h_{i1M} + h_{i2M} VDUM_{im}] HML_m + p_{iM} PR1YR_t + e_{im} \\ m &= 1, 2, ..., M \end{aligned} \tag{5}$$

$SDUM_{im}$ is the percentage of small cap funds minus the percentage of large cap funds in portfolio i at time m, while $VDUM_{im}$ is the percentage of value funds minus the percentage of growth funds in the portfolio i at time m. Since the factor loading on SMB varies with the size of the portfolio, the interactive term between SDUM and SMB captures the variability of SMB factor loadings through time. Also since the factor loading on HML varies with the book-to-market equity of the portfolio, the interactive term between VDUM and HML captures the variability of the HML factor loadings through time. The results for the conditional regressions, displayed in Tables 1a and 2a, tell a story similar to that told by their unconditional counterparts in Tables 1 and 2.[16]

Overall the results suggest that style-adjusted abnormal return differences persist for up to six years while non-style-adjusted abnormal return differences do not. The results thus far seem to

[16] In results not reported, we perform the standard and conditional 3-factor and 4-factor regressions on style-adjusted returns with expenses added back in. We do so in response to the critique that persistence in expense ratios drives most of the persistence in fund returns. We find the 3-factor and 4-factor spreads to be significantly positive up to and including the sort on style-adjusted returns four to six years ago. For the four- to six-year style-adjusted return sort, the 3-factor spread is 2.4% per year at the 1% significance level while the 4-factor spread is 2.5% at the 1% significance level. Only the spreads from the sort on style-adjusted returns five to seven years ago lose their significance when expenses are added back in. These results are available upon request.

Table 1a

Portfolios of Mutual Funds Formed on Three-Year Sums of Past Style-Adjusted Fund Returns

Mutual funds are sorted on January 1 each year from January 1984 to December 1999 into ten portfolios based on the three-year sum of their returns minus their styles' returns. The portfolios are equally weighted monthly so the weights are re-adjusted whenever a stock disappears. Spread A is the return of the funds in decile 1 minus the return of the funds in decile 10. Spread B is the return of the funds in decile 2 minus the return of the funds in decile 10. Spread C is the return of the funds in decile 1 minus the returns of the funds in decile 9. 123spreadA denotes the spread A derived from a sort on sum of past one-, two-, and three-year style-adjusted returns. Similarly 234spreadA denotes the spread A derived from a sort on the sum of past two-, three-, and four-year style-adjusted returns. SDUM is the percentage of small cap funds minus the percentage of large cap funds in the portfolio. VDUM is the percentage of value funds minus the percentage of growth funds in the portfolio. RMRF, SMB, and HML are Fama and French's (1993) market proxy and factor-mimicking portfolios for size and book-to-market equity. PR1YR is Carhart's (1997) factor-mimicking portfolio for one-year return momentum. Alpha is the intercept of the Model. The t-statistics are in parentheses. The number of observations for each regression is 192.

	Conditional Fama-French 3-factor Model							Conditional Carhart 4-factor Model							
Portfolio	Alpha	RMRF	SMB	SDUM*SMB	HML	VDUM*HML	Adj R-sq	Alpha	RMRF	SMB	SDUM*SMB	HML	VDUM*HML	PR1YR	Adj R-sq
123spreadA	**0.40**	-0.02	-0.09	-0.10	-0.15	0.25	0.061	**0.34**	-0.01	-0.05	-0.04	-0.14	0.29	0.06	0.087
	(4.83)**	(-0.72)	(-1.85)	(-0.43)	(-3.75)	(1.42)		**(3.94)****	(-0.65)	(-1.03)	(-0.19)	(-3.45)	(1.65)	(2.51)	
123spreadB	**0.35**	-0.01	-0.16	-0.16	-0.08	0.03	0.070	**0.30**	-0.01	-0.13	-0.11	-0.08	0.08	0.05	0.089
	(5.13)**	(-0.65)	(-2.49)	(-0.98)	(-1.68)	(0.26)		**(4.26)****	(-0.62)	(-1.84)	(-0.68)	(-1.75)	(0.64)	(2.24)	
123spreadC	**0.20**	-0.02	-0.03	-0.01	-0.13	-0.28	0.104	**0.17**	-0.02	-0.02	-0.04	-0.13	-0.28	0.03	0.109
	(3.16)**	(-1.30)	(-1.05)	(-0.03)	(-4.92)	(-1.99)		**(2.54)***	(-1.27)	(-0.68)	(-0.23)	(-4.61)	(-2.03)	(1.47)	
234spreadA	**0.30**	-0.02	-0.02	0.29	-0.16	0.44	0.095	**0.29**	-0.02	-0.02	0.29	-0.16	0.45	0.01	0.090
	(4.09)**	(-1.06)	(-0.56)	(1.03)	(-4.53)	(2.91)		**(3.78)****	(-1.05)	(-0.48)	(1.02)	(-4.43)	(2.91)	(0.34)	
234spreadB	**0.33**	-0.03	-0.13	-0.07	-0.13	0.21	0.090	**0.32**	-0.03	-0.13	-0.07	-0.13	0.21	0.01	0.086
	(4.94)**	(-1.52)	(-1.89)	(-0.41)	(-2.74)	(1.63)		**(4.59)****	(-1.51)	(-1.83)	(-0.40)	(-2.67)	(1.62)	(0.34)	
234spreadC	0.09	0.03	0.01	0.34	-0.09	0.06	0.189	0.10	0.03	0.01	0.34	-0.09	0.05	-0.01	0.186
	(1.77)	(2.11)	(0.51)	(2.55)	(-4.32)	(0.69)		(1.85)	(2.08)	(0.30)	(2.50)	(-4.34)	(0.63)	(-0.55)	
345spreadA	**0.22**	0.02	-0.06	-0.15	-0.10	0.01	0.104	**0.22**	0.02	-0.05	-0.15	-0.10	0.01	0.00	0.100
	(3.55)**	(1.14)	(-1.30)	(-0.65)	(-3.92)	(0.10)		**(3.34)****	(1.14)	(-1.26)	(-0.66)	(-3.80)	(0.07)	(0.14)	
345spreadB	**0.25**	-0.02	-0.09	-0.05	-0.05	-0.08	0.056	**0.25**	-0.02	-0.09	-0.05	-0.05	-0.09	0.00	0.051
	(4.66)**	(-1.15)	(-1.63)	(-0.44)	(-1.83)	(-0.95)		**(4.37)****	(-1.15)	(-1.61)	(-0.44)	(-1.74)	(-0.95)	(0.10)	
345spreadC	0.02	0.03	0.03	0.41	-0.07	0.29	0.238	0.04	0.02	0.02	0.40	-0.07	0.30	-0.02	0.245
	(0.41)	(2.16)	(1.80)	(3.95)	(-3.55)	(3.38)		(0.88)	(2.12)	(1.19)	(3.86)	(-3.79)	(3.49)	(-1.59)	
456spreadA	**0.24**	0.02	-0.05	-0.11	-0.14	0.03	0.159	**0.25**	0.02	-0.06	-0.10	-0.14	0.03	-0.01	0.156
	(3.60)**	(1.19)	(-1.00)	(-0.43)	(-4.74)	(0.27)		**(3.59)****	(1.17)	(-1.06)	(-0.41)	(-4.76)	(0.27)	(-0.55)	
456spreadB	**0.22**	-0.01	0.03	0.25	-0.06	0.01	0.062	**0.24**	-0.01	0.03	0.27	-0.07	-0.00	-0.02	0.067
	(3.89)**	(-0.58)	(0.62)	(2.00)	(-2.10)	(0.19)		**(4.16)****	(-0.64)	(0.54)	(2.13)	(-2.19)	(-0.00)	(-1.46)	
456spreadC	**0.13**	0.04	0.04	0.22	-0.09	0.06	0.241	**0.16**	0.04	0.03	0.23	-0.10	0.07	-0.03	0.248
	(2.54)*	(3.12)	(1.90)	(1.75)	(-3.91)	(0.77)		**(2.95)****	(3.10)	(1.26)	(1.84)	(-4.16)	(0.89)	(-1.67)	
567spreadA	**0.12**	0.04	0.06	0.24	-0.10	0.03	0.236	**0.13**	0.04	0.05	0.24	-0.10	0.03	-0.01	0.234
	(2.21)*	(3.33)	(1.63)	(1.87)	(-4.15)	(0.50)		**(2.32)***	(3.30)	(1.46)	(1.87)	(-4.21)	(0.42)	(-0.73)	
567spreadB	**0.13**	0.01	0.07	0.26	-0.03	0.03	0.055	**0.15**	0.01	0.06	0.25	-0.03	0.00	-0.02	0.062
	(2.36)*	(0.44)	(2.11)	(3.57)	(-0.98)	(0.38)		**(2.70)****	(0.40)	(1.67)	(3.48)	(-1.02)	(0.06)	(-1.54)	
567spreadC	-0.01	0.10	0.05	0.19	-0.08	0.08	0.538	-0.00	0.10	0.05	0.20	-0.08	0.08	-0.01	0.536
	(-0.25)	(9.12)	(2.95)	(2.13)	(-4.14)	(1.72)		(-0.09)	(9.09)	(2.66)	(2.16)	(-4.16)	(1.68)	(-0.49)	

**Alpha significant at the 1% level, *Alpha significant at the 5% level

Table 2a

Portfolios of Mutual Funds Formed on Three-Year Sums of Past Fund Returns

Mutual funds are sorted on January 1 each year from January 1984 to December 1999 into ten portfolios based on three-year sums of their returns. The portfolios are equally weighted monthly so the weights are re-adjusted whenever a stock disappears. Spread A is the return of the funds in decile 1 minus the return of the funds in decile 10. Spread B is the return of the funds in decile 2 minus the return of the funds in decile 10. Spread C is the return of the funds in decile 1 minus the returns of the funds in decile 9. 123spreadA denotes the spread A derived from a sort on sum of past one-, two-, and three-year returns. Similarly 234spreadA denotes the spread A derived from a sort on the sum of past two-, three-, and four-year returns. SDUM is the percentage of small cap funds minus the percentage of large cap funds in the portfolio. VDUM is the percentage of value funds minus the percentage of growth funds in the portfolio. RMRF, SMB, and HML are Fama and French's (1993) market proxy and factor-mimicking portfolios for size and book-to-market equity. PR1YR is Carhart's (1997) factor-mimicking portfolio for one-year return momentum. Alpha is the intercept of the Model. The t-statistics are in parentheses. The number of observations for each regression is 192.

	Conditional Fama-French 3-factor Model							Conditional Carhart 4-factor Model							
Portfolio	Alpha	RMRF	SMB	SDUM*SMB	HML	VDUM*HML	Adj R-sq	Alpha	RMRF	SMB	SDUM*SMB	HML	VDUM*HML	PR1YR	Adj R-sq
123spreadA	**0.51**	0.01	-0.03	0.47	-0.11	0.57	0.533	**0.45**	0.01	-0.01	0.48	-0.10	0.56	0.05	0.540
	(5.52)**	(0.42)	(-0.93)	(9.71)	(-2.69)	(9.25)		**(4.68)****	(0.44)	(-0.22)	(9.95)	(-2.39)	(9.12)	(2.01)	
123spreadB	**0.48**	-0.02	-0.06	0.41	-0.15	0.59	0.574	**0.44**	-0.02	-0.05	0.42	-0.14	0.59	0.04	0.578
	(6.18)**	(-0.98)	(-1.91)	(8.46)	(-4.33)	(9.90)		**(5.42)****	(-0.97)	(-1.28)	(8.62)	(-4.00)	(9.76)	(1.57)	
123spreadC	**0.26**	0.00	0.03	0.48	-0.11	0.46	0.558	**0.23**	0.00	0.05	0.48	-0.10	0.46	0.03	0.559
	(3.37)**	(0.14)	(1.17)	(10.98)	(-2.92)	(8.74)		**(2.84)****	(0.14)	(1.49)	(11.05)	(-2.75)	(8.59)	(1.20)	
234spreadA	**0.40**	0.01	-0.06	0.56	-0.06	0.53	0.517	**0.37**	0.01	-0.04	0.57	-0.05	0.54	0.04	0.521
	(5.00)**	(0.43)	(-1.74)	(8.06)	(-1.70)	(10.65)		**(4.34)****	(0.50)	(-1.17)	(8.16)	(-1.37)	(10.80)	(1.54)	
234spreadB	**0.34**	-0.00	-0.05	0.49	-0.09	0.51	0.601	**0.32**	-0.00	-0.04	0.49	-0.08	0.52	0.02	0.601
	(5.31)**	(-0.08)	(-2.05)	(8.69)	(-2.88)	(11.38)		**(4.74)****	(-0.03)	(-1.53)	(8.76)	(-2.66)	(11.44)	(1.09)	
234spreadC	**0.17**	0.05	-0.01	0.64	-0.04	0.48	0.629	**0.17**	0.05	-0.02	0.64	-0.04	0.48	-0.01	0.627
	(2.56)*	(2.86)	(-0.47)	(10.77)	(-1.32)	(11.74)		**(2.55)***	(2.83)	(-0.56)	(10.75)	(-1.37)	(11.49)	(-0.37)	
345spreadA	**0.27**	0.03	-0.08	0.68	-0.01	0.60	0.615	**0.22**	0.03	-0.06	0.69	-0.00	0.60	0.04	0.622
	(3.94)**	(1.66)	(-2.43)	(9.80)	(-0.41)	(11.21)		**(3.12)****	(1.74)	(-1.88)	(10.00)	(-0.05)	(11.34)	(2.13)	
345spreadB	**0.29**	-0.00	-0.02	0.50	-0.06	0.43	0.550	**0.26**	-0.00	-0.01	0.51	-0.05	0.44	0.02	0.550
	(4.77)**	(-0.29)	(-0.89)	(9.99)	(-2.16)	(9.56)		**(4.20)****	(-0.24)	(-0.45)	(10.06)	(-1.91)	(9.62)	(1.16)	
345spreadC	0.02	0.06	-0.01	0.63	-0.00	0.57	0.718	0.01	0.06	-0.01	0.63	0.00	0.57	0.01	0.717
	(0.37)	(3.92)	(-0.43)	(11.57)	(-0.09)	(12.03)		(0.17)	(3.93)	(-0.26)	(11.56)	(0.03)	(12.02)	(0.59)	
456spreadA	**0.30**	0.04	-0.08	0.66	0.01	0.64	0.637	**0.26**	0.04	-0.06	0.66	0.02	0.66	0.04	0.641
	(4.01)**	(1.96)	(-2.12)	(8.71)	(0.30)	(12.69)		**(3.34)****	(2.05)	(-1.62)	(8.76)	(0.67)	(12.87)	(1.71)	
456spreadB	**0.28**	-0.00	-0.08	0.60	-0.01	0.52	0.594	**0.27**	-0.00	-0.08	0.60	-0.01	0.52	0.00	0.592
	(4.27)**	(-0.12)	(-3.29)	(9.81)	(-0.52)	(12.96)		**(4.03)****	(-0.11)	(-3.07)	(9.78)	(-0.48)	(12.72)	(0.14)	
456spreadC	**0.14**	0.06	-0.01	0.58	0.02	0.65	0.740	0.12	0.06	0.00	0.58	0.02	0.66	0.02	0.741
	(2.26)*	(3.70)	(-0.19)	(9.80)	(0.50)	(12.75)		(1.79)	(3.75)	(0.10)	(9.82)	(0.76)	(12.82)	(1.21)	
567spreadA	**0.14**	0.06	-0.03	0.53	0.05	0.62	0.660	0.11	0.06	-0.01	0.53	0.06	0.64	0.03	0.664
	(2.09)*	(3.22)	(-0.97)	(8.52)	(1.67)	(14.67)		(1.56)	(3.26)	(-0.42)	(8.46)	(2.02)	(14.75)	(1.68)	
567spreadB	0.12	0.03	-0.07	0.52	0.04	0.52	0.601	0.11	0.03	-0.06	0.52	0.04	0.52	0.01	0.599
	(1.86)	(1.94)	(-2.78)	(9.88)	(1.44)	(13.91)		(1.69)	(1.94)	(-2.51)	(9.78)	(1.47)	(13.63)	(0.33)	
567spreadC	-0.02	0.10	0.03	0.48	0.09	0.60	0.766	-0.06	0.10	0.05	0.47	0.11	0.62	0.04	0.772
	(-0.37)	(7.17)	(1.26)	(9.24)	(3.45)	(15.20)		(-1.01)	(7.29)	(1.90)	(9.21)	(3.95)	(15.56)	(2.36)	

**Alpha significant at the 1% level, *Alpha significant at the 5% level suggest that differences in managerial ability are present and are better captured by the style-adjusted return metric.

4.2 Style-Adjusted Returns Versus Fund Expenses

Carhart (1992) suggests that long-term mutual fund performance persistence is fueled by persistence in expenses. In this section, we test this assertion directly by measuring the explanatory power of style-adjusted returns on future style-adjusted alphas after accounting for expense ratios. We estimate Fama-MacBeth (1973) cross-sectional regressions on the monthly 3-factor residuals of funds minus the monthly 3-factor residuals of the funds' styles (style-adjusted fund 3-factor alpha). We include as independent variables: style-adjusted returns, expense ratio, total net assets, and load at various yearly lags. More specifically, we estimate the following regressions each month:

$$SA\alpha_{im} = a_m + b_m SAret_{it-k} + \varepsilon_{im} \quad i = 1, ..., N,\ t = 1, ..., T,\ m = 1, ..., 12T \tag{6}$$

$$\begin{aligned} SA\alpha_{im} &= a_m + b_m SAret_{it-k} + c_m EXP_{it-k} + d_m LOAD_{it-k} + e_m TNA_{it-k} + \varepsilon_{im} \\ i &= 1, ..., N,\ t = 1, ..., T,\ m = 1, ..., 12T \end{aligned} \tag{7}$$

where m is a month in year t, $SA\alpha_{im}$ is style-adjusted α_{im}, α_{im} is an individual fund performance estimate, k is the yearly lag of the independent variables, $SAret$ is the style-adjusted return (after expense) of the fund, EXP is expense ratio, $LOAD$ is load, and TNA is total net assets of the fund. As in Fama and MacBeth (1973), we estimate the cross-sectional regression each month, then average the coefficient estimates across the complete sample period. This yields 192 cross-sectional regressions which average 833 observations each for a combined sample of about 160,000 observations.[17] To mitigate look-ahead bias, we estimate α_{im} as the one-month abnormal return from the 3-factor model, where the 3-factor model loadings are estimated over the prior 36 months[18]:

[17] These numbers are for the multivariate Fama-MacBeth regressions with style-adjusted returns, expense ratio, load, and TNA at a lag of one year as explanatory variables. The number of observations for the multivariate regressions with explanatory variables lagged seven years is smaller at around 352 per regression for a combined total of 67,600 observations.

[18] If a fund does not have 36 months of past return data, we use a minimum of 30 months of past return data to estimate the alphas.

$$\alpha_{im} \equiv R_{im} - R_{Fm} - \hat{b}_{im-1} RMRF_m - \hat{s}_{im-1} SMB_m - \hat{h}_{im-1} HML_m \qquad i = 1, ..., N, \; m = 1, ..., M \tag{8}$$

where $\hat{b}$, $\hat{s}$, and $\hat{h}$ are the estimated loadings. If persistence in expense ratios accounts entirely for the long-term mutual fund performance persistence, then the coefficients on style-adjusted return (when expense ratio is included as an independent variable) should be insignificantly different from zero for lags greater than two to three years.[19]

The univariate regression results in Table 3 suggest that long-term mutual fund performance persistence is strong. Style-adjusted fund return has significant explanatory power on future style-adjusted 3-factor fund alpha even for lags as long as seven years. A one percentage-point increase in style-adjusted returns of a fund seven years ago increases the monthly style-adjusted 3-factor alpha of the fund by 0.41 percentage points. The multivariate regression results in Table 3 indicate that neither expense ratio, load, nor TNA can account for the explanatory power of style-adjusted returns at long lags. The coefficients on style-adjusted fund returns, for lags between four and seven years, fall by an average of only 29% when expense ratio, load, and TNA are accounted for. After accounting for expense ratio, load, and TNA, a one percentage-point increase in style-adjusted returns of a fund seven years ago increases the monthly style-adjusted 3-factor alpha of the fund by 0.25 percentage points. The coefficient estimates for the regressions with non-style-adjusted fund alphas and returns (displayed at the bottom of Table 3) exhibit less persistence than those for the regressions with style-adjusted fund alphas and returns. This is more in line with Carhart's findings. Note however that the coefficient on fund return still has strong and significant explanatory power on fund alpha after four years.

Next, we estimate the above cross-sectional regressions again but this time using as α_{im} the one-month abnormal returns from the Carhart (1997) 4-factor model, where the 4-factor model loadings are estimated over the prior 36 months (or over a minimum of 30 months if the fund has less than 36 months of return data):

$$\alpha_{im} \equiv R_{im} - R_{Fm} - \hat{b}_{im-1} RMRF_m - \hat{s}_{im-1} SMB_m - \hat{h}_{im-1} HML_m - \hat{p}_{im-1} PR1YR_m$$

[19] We are taking long-term performance persistence to be performance persistence for four years and beyond.

Table 3

Fama-MacBeth (1973) Cross-Sectional Regressions on Fama-French 3-Factor Alphas

Cross-sectional univariate and multivariate regressions are estimated for each month from January 1984 to December 1999 across all funds in the sample at that time. The dependent variable is the monthly Fama-French 3-factor residual of the fund minus the monthly 3-factor residual of the fund's style, where the factor loadings are estimated on the prior three years of monthly returns. The independent variables are style-adjusted fund return, expense ratio, fund TNA, fund maximum load fees, all at a certain yearly lag. Style-adjusted fund return in a year is fund return minus style return in that year. Fund return is mean monthly return over the year. Expense ratio is management, administrative, and 12b-1 expenses divided by average TNA. TNA is total net assets. Maximum load is the sum of maximum front-end, back-end, and deferred sales charges. In each univariate regression, the independent variable is a style-adjusted fund return. In each multivariate regression, the independent variables are style-adjusted fund return, expense ratio, TNA, and load. The reported estimates are the time-series averages of monthly cross-sectional regression slope estimates as in Fama and MacBeth (1973). For the univariate regression with the independent variable lagged one year, a total of 192 cross-sectional regressions are estimated which average 883 observations each for a combined sample of about 169,500 observations. For the multivariate regression with the independent variables lagged one year, a total of 192 cross-sectional regressions are estimated which average 833 observations each for a combined sample of about 159,900 observations. The t-statistics, in parentheses, are on the time-series means of the coefficients. The coefficients for the fund attributes are suppressed for brevity.

Independent Variables (Coefficients x 100)	Univariate regressions	Multivariate regressions (controlling for fund attributes)
Style-adjusted fund return (t-1)	151.79	138.70
	(6.17)***	(5.35)***
Style-adjusted fund return (t-2)	92.80	73.94
	(4.31)***	(3.18)***
Style-adjusted fund return (t-3)	40.81	26.19
	(2.08)**	(1.24)
Style-adjusted fund return (t-4)	63.90	53.50
	(3.99)***	(3.19)***
Style-adjusted fund return (t-5)	47.67	33.84
	(3.19)***	(2.23)**
Style-adjusted fund return (t-6)	56.85	39.13
	(3.66)***	(2.66)***
Style-adjusted fund return (t-7)	41.06	24.64
	(3.37)***	(2.11)**
Fund return (t-1)	137.90	126.85
	(4.90)***	(4.31)***
Fund return (t-2)	78.05	65.63
	(3.07)***	(2.46)**
Fund return (t-3)	36.61	27.29
	(1.63)	(1.15)
Fund return (t-4)	74.50	69.39
	(3.74)***	(3.37)***
Fund return (t-5)	34.70	27.02
	(2.02)**	(1.52)
Fund return (t-6)	35.12	25.00
	(1.70)*	(1.32)
Fund return (t-7)	34.66	21.41
	(1.91)*	(1.19)

***Significant at the 1 % level
**Significant at the 5% level
*Significant at the 10 % level

Table 4

Fama-MacBeth (1973) Cross-Sectional Regressions on Carhart 4-Factor Alphas

Cross-sectional univariate and multivariate regressions are estimated for each month from January 1984 to December 1999 across all funds in the sample at that time. The dependent variable is the monthly Carhart 4-factor residual of the fund minus the monthly 4-factor residual of the fund's style, where the factor loadings are estimated on the prior three years of monthly returns. The independent variables are style-adjusted fund return, expense ratio, fund TNA, fund maximum load fees, all at a certain yearly lag. Style-adjusted fund return in a year is fund return minus style return in that year. Fund return is mean monthly return over the year. Expense ratio is management, administrative, and 12b-1 expenses divided by average TNA. TNA is total net assets. Maximum load is the sum of maximum front-end, back-end, and deferred sales charges. In each univariate regression, the independent variable is a style-adjusted fund return. In each multivariate regression, the independent variables are style-adjusted fund return, expense ratio, TNA, and load. The reported estimates are the time-series averages of monthly cross-sectional regression slope estimates as in Fama and MacBeth (1973). For the univariate regression with the independent variable lagged one year, a total of 192 cross-sectional regressions are estimated which average 883 observations each for a combined sample of about 169,500 observations. For the multivariate regression with the independent variables lagged one year, a total of 192 cross-sectional regressions are estimated which average 833 observations each for a combined sample of about 159,900 observations. The t-statistics, in parentheses, are on the time-series means of the coefficients. The coefficients for the fund attributes are suppressed for brevity.

Independent Variables (Coefficients x 100)	Univariate regressions	Multivariate regressions (controlling for fund attributes)
Style adjusted fund return (t-1)	144.42	135.27
	(6.03)***	(5.33)***
Style adjusted fund return (t-2)	59.71	43.84
	(2.57)**	(1.76)*
Style adjusted fund return (t-3)	3.44	-12.30
	(0.17)	(-0.56)
Style adjusted fund return (t-4)	52.75	39.36
	(3.22)***	(2.33)**
Style adjusted fund return (t-5)	47.67	39.28
	(3.19)***	(2.65)***
Style adjusted fund return (t-6)	48.79	40.14
	(3.11)***	(2.64)***
Style adjusted fund return (t-7)	44.09	21.33
	(3.57)***	(1.80)*
Fund return (t-1)	129.03	120.84
	(4.71)***	(4.22)***
Fund return (t-2)	44.58	33.49
	(1.72)*	(1.22)
Fund return (t-3)	-0.44	-12.30
	(-0.02)	(-0.52)
Fund return (t-4)	50.84	41.96
	(2.59)**	(2.08)**
Fund return (t-5)	31.64	22.92
	(1.84)*	(1.24)
Fund return (t-6)	30.46	21.97
	(1.56)	(1.13)
Fund return (t-7)	36.31	17.30
	(2.10)**	(1.00)

***Significant at the 1 % level

**Significant at the 5% level

*Significant at the 10 % level

$$i = 1,...,N,\ m = 1,...,M \tag{9}$$

where $\hat{b}$, $\hat{s}$, $\hat{h}$, and $\hat{p}$ are the estimated loadings. The results displayed in Table 4 are very similar to the results in Table 3. We find that style-adjusted returns up to the seventh yearly lag still have strong and significant explanatory power on style-adjusted 4-factor alpha. After accounting for lagged expense ratios, load, and TNA,[20] style-adjusted returns up to the sixth yearly lag significantly (at the 5% level) explain style-adjusted 4-factor alpha. A one percentage-point increase in style-adjusted returns six years ago increases style-adjusted 4-factor alpha by 0.40 percentage points.

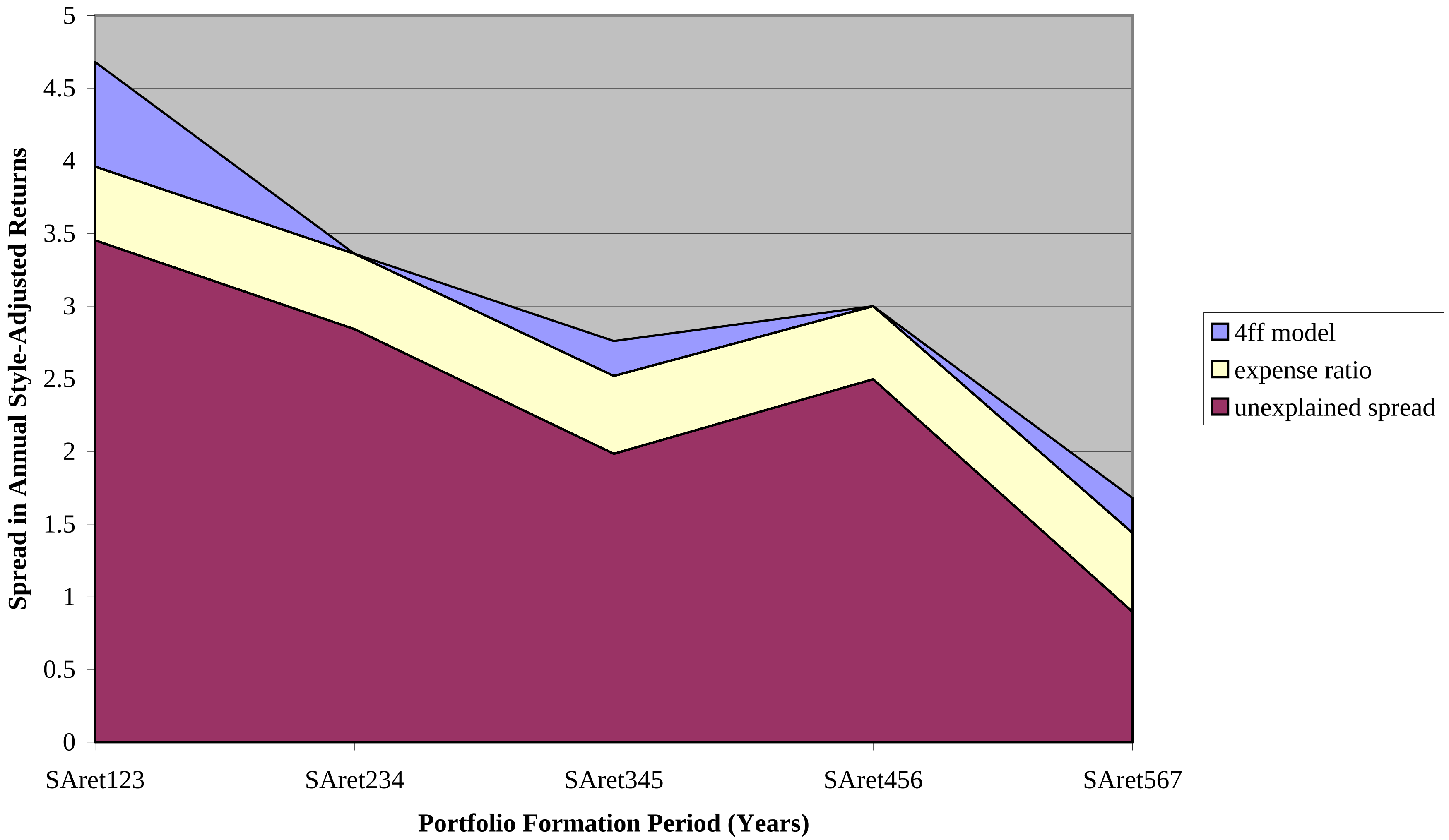
4.3 More About Expenses

To further investigate the purported prowess of expenses in explaining long-term persistence, we find the yearly differences in mean style-adjusted expense ratios and turnovers[21] between funds in the top and bottom deciles of the sorts in Table 1. For each sort, we calculate the t-statistics that test the hypothesis that the means of the differences are zero. The style-adjusted expense ratios of the bottom decile funds are indeed consistently higher than those of the top decile funds. Moreover, the expense ratio spreads are highly significant and consistent across the different sorts at around 0.52% per year. Contrary to Carhart's findings, however, the turnover spreads are either positively significant or insignificantly different from zero. The turnover spread for the sort on style-adjusted returns one to three years ago is 12% per year at the 5% significance level, while three of the remaining four turnover spreads are positive as well (but insignificantly different from zero). In results not reported, we estimate Fama-MacBeth cross-sectional univariate regressions on 3-factor style-adjusted fund alphas and on 4-factor style-adjusted fund alphas, using contemporaneous turnover as the explanatory variable. We find that the coefficient on turnover is positive at the 10% level for the 3-factor alpha regression and insignificant for the 4-factor alpha regression. Moreover,

[20] It is inconceivable that lagged turnover will have any direct influence on future style-adjusted alpha other than through its association with lagged style-adjusted return. Consistent with this, when we estimate univariate regressions on 3-factor style-adjusted alpha with lagged turnover as the explanatory variable, the coefficient on lagged turnver insignificantly differs from zero. Hence we do not include lagged turnover as a control variable.

[21] Following Carhart, we use as our turnover measure reported turnover plus one-half of the change in fund TNA adjusted for returns and mergers. We do so as reported turnover is the *minimum* of sales and purchases over TNA.

Figure 5
Summary of possible explanations for persistence in mutual fund performance



when we estimate pooled OLS regressions on the alphas with contemporaneous turnover and yearly dummies, contemporaneous turnover is significantly positive at the 1% level (with White (1980) standard errors). Hence our results suggest that higher turnover does somewhat improve a fund's performance relative to the other funds in its style. This finding bodes well for aficionados of active fund management.

To illustrate the explanatory power of the 4-factor model and expense ratio in explaining the spread in mean excess returns, we plot in Figure 5 the spread in style-adjusted returns that is explained either by expense ratio or the 4-factor model. Clearly, expense ratio and the 4-factor model do not explain much of the spread in style-adjusted returns. The 4-factor model and expense ratios combined only account for, on average, less than 25% of the spreads for the various sorts. Moreover, since the turnover spreads are either positive or insignificantly different from zero, differences in transactions cost cannot account for any of the "unexplained spread" in Figure 5. This contrasts sharply with Carhart's finding that the 4-factor model, expense ratios, and transactions cost account for almost two-thirds of the spread in mean returns when funds are sorted on their one-year, one- to two-year, one- to three-year, and one- to four-year returns. One reason for this difference is that the 4-factor model loses almost all its explanatory power once past one-year returns are taken out of the formation period.

4.4 Consistency in Rankings

In this section we test whether the rankings made on three-year style-adjusted returns are consistent by constructing a contingency table of initial and subsequent three-year style-adjusted fund return rankings. We use simple returns gross of expense ratios to remove the predictable expense element in reported returns and to be consistent with Carhart (1997).

The contingency table is represented graphically in Figure 6a. Three salient results quickly jump out at the reader. First, the funds in the bottom decile (initial ranking equals to ten) tend to remain in the bottom decile (subsequent ranking equals to ten). The probability that a fund in the bottom decile remains in the bottom decile conditional upon surviving is 0.24. Next, the rate of death increases almost monotonically with the rank of the funds. In other words, funds in the lower deciles (with lower past style-adjusted returns) have higher chances of dying. Finally, the funds in the top decile have a somewhat higher probability of remaining in the top decile. The

Figure 6a
Contingency table of initial and subsequent three-year style-adjusted return rankings

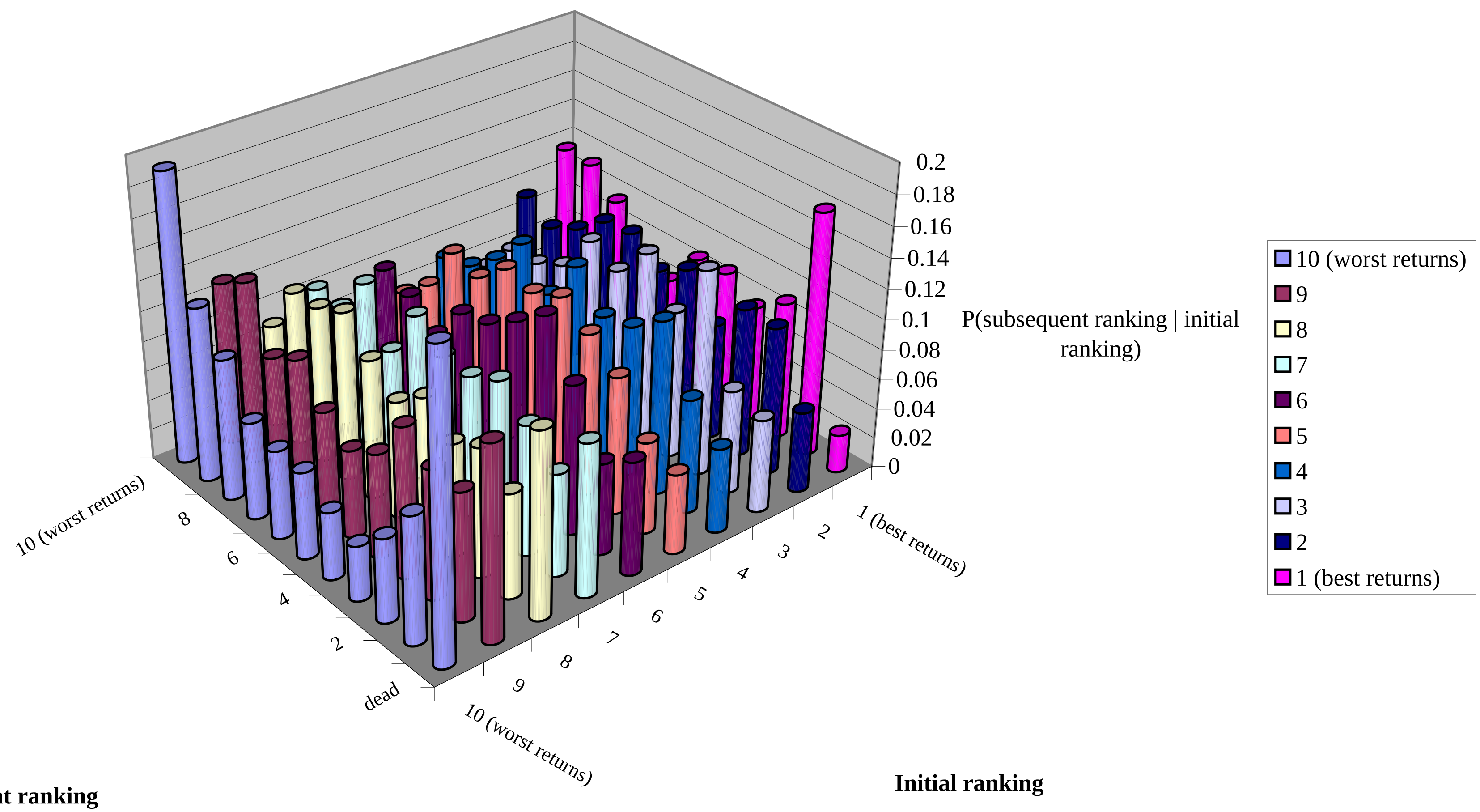
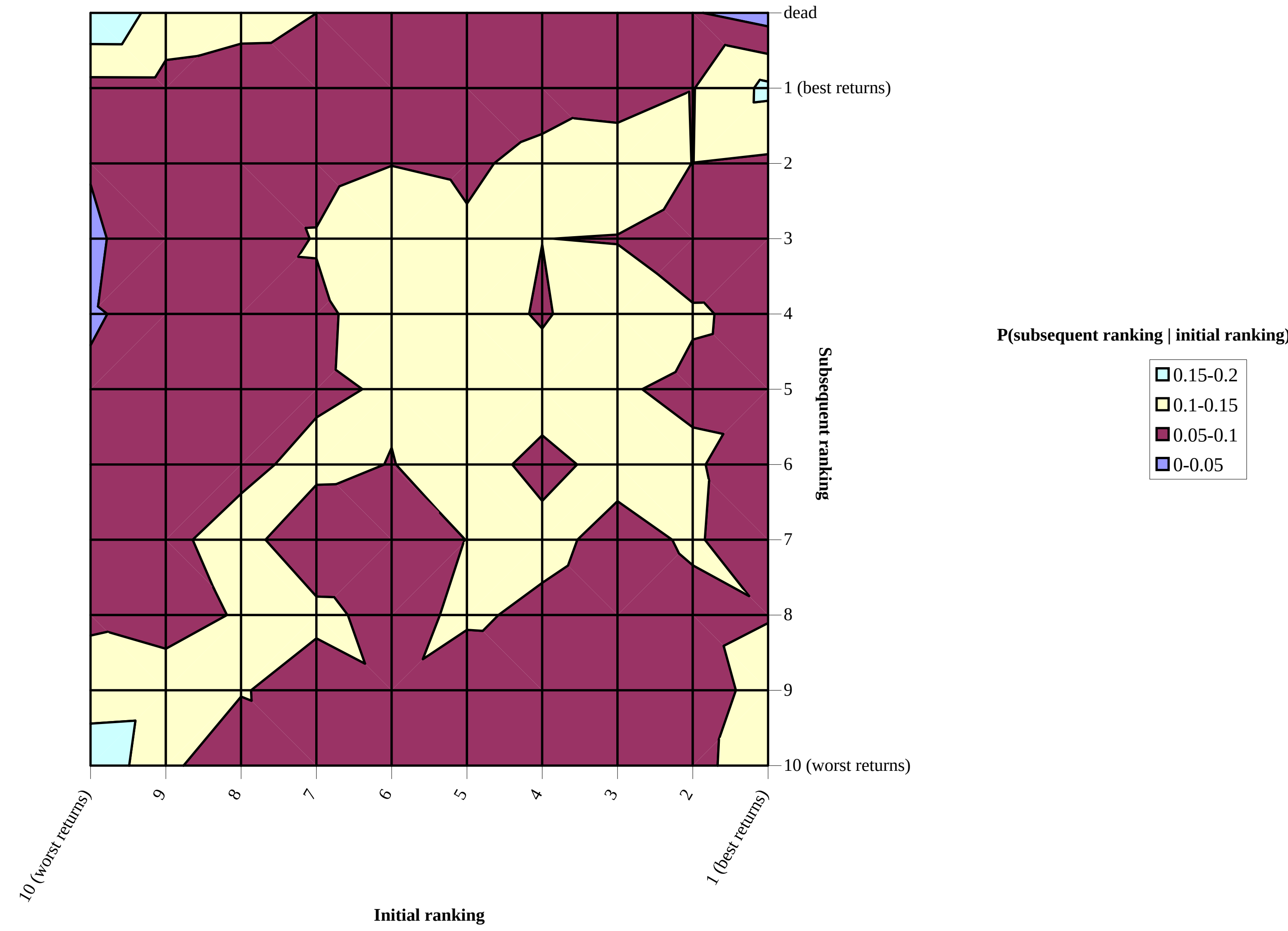


Figure 6b
Contingency contour plot of intial and subsequent three-year style-adjusted return rankings



results for the funds in the less extreme deciles are less clear from Figure 4a.

To help characterize the conditional probabilities for funds in the less extreme deciles, we plot a contour map[22] of the conditional probabilities as shown in Figure 6b. The unshaded diagonal ridge-like region in the contour map suggests that funds in decile x are also somewhat likely to be in decile $x-1$, x, or $x+1$ subsequently. While Figures 6a and 6b do not prove that the rankings are absolutely consistent, they nonetheless show that the rankings are somewhat consistent and are not entirely artifacts of chance.

To better investigate the consistency properties of style-adjusted returns, we now bring statistical inference machinery to bear on the style-adjusted sort. In a nonparametric approach, we construct a contingency table of winners and losers[23]. A fund that beats the median fund (for a certain performance metric) in any period is labeled a winner, and one that doesn't is labeled a loser. We compare a fund's performance in the current period (where a period is defined as one year or three years) to its performance in the previous period. Hence, persistence in this context refers to the existence of funds that are winners in two consecutive periods (denoted by WW) or losers in two consecutive periods (denoted by LL). Letting WL denote winners in the first period and losers in the second period, and LW denote the reverse, we can calculate the cross-product ratio (CPR), which is defined as $(WW * LL)/(WL * LW)$. The CPR ratio captures the ratio of the funds that show persistence in performance to those that do not. Under the hypothesis of no managerial skill, which implies no persistence in gross returns, the probability of winning or losing in each period equals one-half and is independent of the return horizons. So one would expect that the four categories WW, WL, LW, and LL each have 25% of the funds, and CPR equals 1. Since the standard error of the natural logarithm of the CPR is given by

$$\sigma_{\ln(CPR)} = \sqrt{\frac{1}{WW} + \frac{1}{WL} + \frac{1}{LW} + \frac{1}{LL}} \tag{10}$$

(see Christensen, 1990), we can test for the statistical significance of the CPR.

We also conduct a chi-square test comparing the observed frequency distribution of WW, WL, LW, and LL with the expected frequency distribution. Carpenter and Lynch (1999) study the

[22] The contour map is a bird's-eye view of the three-dimensional surface connecting the tops of the cylinders in Figure 4a.

[23] All frequencies and probabilities in this framework are conditional on fund survival.

Table 5

Two-Period Performance Persistence of Mutual Funds

All results in this table are based on returns with expenses added back in. Each panel displays the number of occurrences of WW, WL, LW, and LL in the period 1984-1999, the chi-square statistic, and the cross-product ratio (CPR). The statistical significance of the CPR is tested with a Z-statistic, which measures the ratio of the natural log of CPR to its standard error. Panel A is based on style-adjusted returns, while Panel B is based on non-style-adjusted returns. Each panel has two parts, where each part corresponds to a certain definition of a period (either one year or three years). WW denotes funds that are winners in two consecutive periods; LL denotes funds that are losers in two consecutive periods; WL denotes funds that are winners in the first period and losers in the second period; and LW denotes the reverse.

Panel A: Based on style-adjusted returns

One period=one year			One period=three years		
	# of occurrences	frequency of occurrences		# of occurrences	frequency of occurrences
WW	4930	0.28392	WW	2155	0.27522
WL	3907	0.22501	WL	1922	0.24547
LW	3760	0.21654	LW	1736	0.22171
LL	4767	0.27453	LL	2017	0.25760
Total	17364	1	Total	7830	1
chi-square	237.33**		chi-square	34.06**	
CPR	1.5998^^	(Z-stat=15.37)	CPR	1.3027^^	(Z-stat=5.83)

Panel B: Based on non-style-adjusted returns

One period=one year			One period=three years		
	# of occurrences	frequency of occurrences		# of occurrences	frequency of occurrences
WW	5060	0.29141	WW	2083	0.26603
WL	3773	0.21729	WL	2015	0.25734
LW	3652	0.21032	LW	1834	0.23423
LL	4879	0.28098	LL	1898	0.24240
Total	17364	1	Total	7830	1
chi-square	363.80**		chi-square	2.22	
CPR	1.7917^^	(Z-stat=19.01)	CPR	1.0698	(Z-stat=1.49)

** Significant at the 1% level (chi-square critical value=6.63)

^^ Signifcant at the 1% level (t-statistic critical value=2.58) specification and power of various persistence tests and find that the chi-square test based on the number of winners and losers is well specified, powerful, and more robust to the presence of survivorship bias compared to other test methodologies. We compute the chi-square statistic as

$$\frac{(WW - D_1)^2}{D_1} + \frac{(WL - D_2)^2}{D_2} + \frac{(LW - D_3)^2}{D_3} + \frac{(LL - D_4)^2}{D_4} \text{ where} \tag{11}$$

$$\begin{aligned} D_1 &= (WW + WL) * (WW + LW)/N, \\ D_2 &= (WW + WL) * (WL + LL)/N, \\ D_3 &= (LW + LL) * (WW + LW)/N, \\ D_4 &= (LW + LL) * (WL + LL)/N, \text{ and} \\ N &= WW + WL + LW + LL, \end{aligned}$$

and test this statistic at the 5% significance level, which corresponds to a critical value of 3.84 (one degree of freedom).[24]

We conduct the analysis described above using both style-adjusted fund returns and non-style-adjusted fund returns. Table 5 summarizes the results from this nonparametric approach. Panel A, which is based on style-adjusted returns, shows the existence of statistically significant persistence both for two consecutive years and for two consecutive three-year periods. The chi-square statistics and the Z-statistics for the CPR are easily above their 1% critical values. When we investigate non-style-adjusted returns, however, persistence for two consecutive three-year periods is insignificant both with the Z-statistic and the chi-quare statistic. Clearly the consistency properties of the three-year style-adjusted return metric surpasses that of the three-year non-style-adjusted return metric. Hence, fund rankings based on style-adjusted returns are highly persistent, at least for a coarse sort.[25]

[24] Agarwal and Naik (2000) use these tests as well to investigate hedge fund performance persistence.

[25] We also examine the persistence of three-year performance rankings year by year (i.e., choosing as the start of the evaluation period 1984, 1985, etc.). For style-adjusted gross returns, eight out of the fourteen tests on the CPR (since the last evaluation period starts in 1997) yield significant persistence in rankings, while one of the tests yields a reversal in rankings. For non-style-adjusted gross returns, five out of the fourteen tests reveal significant persistence, while two of the tests reveal reversals in rankings. This is consistent with style-adjusted returns being a less noisy indicator of managerial ability.

4.5 Style-Adjusted Returns and Style Timing

Having performed the sorts with style-adjusted returns, we are now in a position to combine that sort with an attempt at style timing. We aim to find the best funds in the best performing style and the worst funds in the worst performing style. Teo and Woo (2001) show that by sorting stocks on the sum of their styles' past four years of flows[26], one can generate a risk-adjusted return spread of between 6.7% to 9.7% per year. Stocks in the least popular (worst flows) style significantly outperform stocks in the most popular (best flows) style. One plausible reason for this result is that investors overreact at the style level (a la De Bondt and Thaler, 1984) and drive the prices of the stocks in the least popular styles below their fundamental values. Hence in this section, we perform the following two-pass sort on the funds in our sample. First, we sort funds based on the sum of their styles' past four years of flows to separate styles with the best and worst returns. Next, we sort funds on their past three-year style-adjusted returns to separate funds with the best and worst style-adjusted returns. Finally, we measure the performance (return) of the resulting portfolios relative to the 3-factor model and the 4-factor model.

A subset of the results displayed in Table 6 lends to a number of salient observations. First, the 3-factor alpha spreads between the portfolio with the best past style-adjusted returns (portfolio 1) and that with the worst past style-adjusted returns (portfolio 10) are significant. This is true for funds in styles with the worst past flows (portfolio A) and for funds in styles with the best past flows (portfolio B). Second, the 3-factor alpha of the portfolio with the best past style-adjusted returns and the worst past style flows (portfolio 1A) is significantly positive. Portfolio 1A earns a return of 31 basis points (3.7% per year) after adjusting for the Fama-French risk factors. Third, the 3-factor alpha of the portfolio with the worst past style-adjusted returns and the best past style flows (portfolio 10B) is significantly negative. Portfolio 10B registers a return of –51 basis points (-6.1% per year) after adjusting for the Fama-French risk factors. Fourth, the 3-factor and 4-factor alpha spreads between portfolio 1A and portfolio 10B are significantly positive. After accounting for the Fama-French risk factors, the spread between 1A and 10B is an impressive 83 basis points (10.0% per year). The momentum factor only accounts for about one-fourth of this spread. The 4-factor alpha for the spread is 59 basis points (7.1% per year).

[26]Style flows is the sum of fund flows in a style normalized by the total market capitaliztion of the style. See Teo and Woo (2001) for the calculation of style market capitalization.

Table 6

Portfolios of Mutual Funds Formed on the Past Three-Year Style-Adjusted Fund Returns and Past Four-Year Style Flows

Mutual funds are sorted on January 1 each year from January 1984 to December 1999 into ninety portfolios based first on their style's past four-year flows and then on their past three-year style-adjusted returns. The portfolios are equally weighted monthly so the weights are re-adjusted whenever a stock disappears. The portfolios with the lowest style flows are denoted with the letter A. The portfolios with the highest style flows are denoted with the letter B. The portfolios with the highest style adjusted returns are denoted with the number 1, while those with the lowest style-adjusted returns are denoted with the number 10. VWRF is the excess return on the CRSP value-weighted market proxy. RMRF, SMB, and HML are Fama and French's (1993) market proxy and factor-mimicking portfolios for size and book-to-market equity. PR1YR is Carhart's (1997) factor-mimicking portfolio for one-year return momentum. Alpha is the intercept of the Model. The t-statistics are in parentheses. The number of observations for each regression is 192.

			Fama-French 3-Factor Model					Carhart 4-Factor Model					
Portfolio	Monthly Excess Return	Std Dev	Alpha	RMRF	SMB	HML	Adj R-sq	Alpha	RMRF	SMB	HML	PR1YR	Adj R-sq
1A (best	1.18%	5.07%	**0.31**	0.98	0.07	-0.32	0.884	0.26	0.99	0.09	-0.31	0.05	0.884
returns)			**(2.42)***	(29.22)	(1.38)	(-5.78)		(1.92)	(29.29)	(1.71)	(-5.48)	(1.25)	
2A	1.03%	5.00%	0.15	1.01	0.10	-0.23	0.897	0.07	1.01	0.13	-0.21	0.07	0.899
			(1.21)	(32.32)	(2.08)	(-4.48)		(0.57)	(32.59)	(2.61)	(-4.10)	(1.93)	
3A	1.02%	4.64%	0.16	0.97	0.08	-0.18	0.927	0.08	0.97	0.11	-0.16	0.07	0.929
			(1.71)	(39.54)	(2.25)	(-4.37)		(0.84)	(40.22)	(3.03)	(-3.90)	(2.66)	
10A (worst	0.80%	4.90%	-0.06	1.00	0.20	-0.19	0.928	-0.12	1.00	0.22	-0.18	0.05	0.928
returns)			(-0.65)	(39.03)	(5.29)	(-4.55)		(-1.11)	(39.22)	(5.54)	(-4.22)	(1.61)	
1A-10A	0.38%	1.44%	**0.38**	-0.02	-0.13	-0.13	0.060	**0.38**	-0.02	-0.13	-0.13	0.00	0.055
			(3.61)**	(-0.60)	(-3.27)	(-2.86)		**(3.41)****	(-0.60)	(-3.07)	(-2.80)	(0.04)	
1B (best	0.69%	5.40%	-0.02	0.95	0.73	-0.08	0.836	-0.02	0.95	0.73	-0.08	0.00	0.835
returns)			(-0.10)	(22.29)	(11.67)	(-1.20)		(-0.11)	(22.22)	(11.02)	(-1.17)	(0.05)	
8B	0.54%	4.90%	-0.16	0.92	0.74	0.12	0.861	-0.15	0.92	0.73	0.12	-0.00	0.860
			(-1.15)	(25.96)	(14.01)	(2.08)		(-1.06)	(25.88)	(13.18)	(2.03)	(-0.10)	
9B	0.55%	5.30%	-0.23	1.02	0.73	0.10	0.864	-0.21	1.01	0.72	0.09	-0.02	0.864
			(-1.59)	(26.71)	(13.04)	(1.59)		(-1.35)	(26.63)	(12.13)	(1.47)	(-0.52)	
10B (worst	0.27%	5.65%	**-0.51**	1.04	0.81	0.07	0.840	-0.33	1.04	0.74	0.03	-0.17	0.849
returns)			**(-3.03)****	(23.61)	(12.51)	(0.98)		(-1.90)	(24.21)	(11.04)	(0.38)	(-3.48)	
1B-10B	0.42%	2.48%	**0.50**	-0.09	-0.08	-0.16	0.014	0.31	-0.09	-0.00	-0.11	0.17	0.062
			(2.70)**	(-1.91)	(-1.13)	(-1.96)		(1.64)	(-1.87)	(-0.05)	(-1.41)	(3.23)	
1A-10B	0.91%	3.65%	**0.83**	-0.06	-0.74	-0.40	0.266	**0.59**	-0.05	-0.65	-0.34	0.22	0.301
			(3.53)**	(-0.91)	(-8.30)	(-3.91)		**(2.46)***	(-0.86)	(-6.99)	(-3.38)	(3.22)	
2A-10B	0.76%	3.59%	**0.66**	-0.03	-0.72	-0.30	0.245	0.40	-0.03	-0.61	-0.24	0.24	0.289
			(2.81)**	(-0.53)	(-8.00)	(-3.00)		(1.67)	(-0.46)	(-6.64)	(-2.43)	(3.53)	
1A-9B	0.63%	3.23%	**0.55**	-0.03	-0.66	-0.42	0.285	**0.47**	-0.03	-0.63	-0.41	0.07	0.287
			(2.68)**	(-0.58)	(-8.46)	(-4.80)		**(2.19)***	(-0.55)	(-7.62)	(-4.53)	(1.16)	

**Alpha significant at the 1% level
*Alpha significant at the 5% level

By combining a sort on past style-adjusted returns and a sort on past style flows, we have been able to separate funds on two dimensions: fund performance and style performance. In doing so, we have levered up the spread between the best and the worst funds by a factor of two relative to that for the sorts on past style-adjusted fund returns. Fund investors may find this a viable strategy for improving the returns on their mutual fund portfolios.

5 Conclusion

The results in this paper have done much to forward the idea that managers are responsible for the persistence in mutual fund performance.[27] We are the first, after Carhart (1997), to use the common-factor methods to re-examine long-term mutual fund performance persistence. Our common-factor regressions are also modified to include time-varying risk factors. The results from these regressions suggest that differences in mutual fund performance persist for up to six years. The momentum factor in the Carhart (1997) 4-factor model does not explain this persistence. By our estimates, persistence in fund expense ratios only explains less than 25% of the differences. Also, fund turnover has a positive effect on fund risk-adjusted performance, indicating that active management is not as detrimental to fund returns as Carhart (1997) suggests. Finally, our style-adjusted methodology opens up the possibility for style-timing. Investors can combine knowledge of funds' past style-adjusted returns with style information to improve the returns on their fund

[27] One concern is that our results may be driven by the timing of stock returns. If what we truly observe is managerial ability, then style-adjusted abnormal return should not just be predictable with past style-adjusted abnormal return but should also be predictable with future style-adjusted abnormal return. To test this, we adopt the methodology of Grinblatt and Titman (1992), and see if style-adjusted abnormal return can still be predicted by "past" style-adjusted abnormal returns after randomizing the months. We focus on funds with 3-factor alpha information for all the months in the last ten years of our sample, since there are more funds in that subsample and Grinblatt and Titman also focus on a ten-year period. See Grinblatt and Titman (1992) for a detailed description of the procedure. We find that the results for the randomized partitioning are even more striking than that for those for the chronological partitioning. For the chronological partitioning, a one percentage-point increase in average style-adjusted alpha in the past sixty months predicts a 0.26 percentage-point increase in style-adjusted alpha in each of the next sixty months (t-statistic = 3.83). For the randomized partitioning (estimating the regression 100 times with different random draws), a one percentage-point increase in average style-adjusted alpha in the "past" 60 months predicts on average a 0.41 percentage-point increase in style-adjusted alpha in each of the "following" sixty months (average t-statistic = 4.78).

portfolios. By sorting on past style flows as well, we have been able to generate a 3-factor-adjusted spread of 10% per year. Our findings nicely complement the results of Wermers (2000) and Chen, Jegadeesh, and Wermers (2000) who use stock holding data to show that mutual fund managers have stock-picking ability.

6 References

Agarwal, Vikas, and Narayan Naik, 2000, Multi-period performance persistence analysis of hedge funds, *Journal of Financial and Quantitative Analysis* 35, 327-342.

Brown, Stephen, and William Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679-698.

Carhart, Mark M., 1992, Persistence in mutual fund performance re-examined, Working paper, Graduate School of Business, University of Chicago.

Carhart, Mark M., 1997, On persistence of mutual fund performance, *Journal of Finance*, 52, 57-82.

Carpenter, Jennifer and Anthony Lynch, 1999, Survivorship bias and attrition effects in measures of performance persistence, *Journal of Financial Economics*, 54, 337-374.

Chen, Hsiu-Lang, Narisimhan Jegadeesh, and Russ Wermers, 2000, The value of active mutual fund management: An examination of the stockholdings and trades of fund managers, *Journal of Financial and Quantitative Analysis.*

Chevalier, Judith and Glenn Ellison, 1996, Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance, Working paper 5852, NBER.

Christensen, Ronald, 1990, *Log-linear models*, Springer-Verlag, New York.

Elton, Edwin J., Martin J. Gruber, and Christopher Blake, 1996, Survivorship bias and mutual fund performance, *The Review of Financial Studies* 9, 1097-1120.

Elton, Edwin J., Martin J. Gruber, and Christopher Blake, 2001, A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases, forthcoming in *Journal of Finance*.

Elton, Edwin J., Martin J. Gruber, and Sanjiv Das, and Christopher Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133-157.

Elton, Edwin J., Martin J. Gruber, and Sanjiv Das, and Matt Hlavka, 1993, Efficiency with costly information: A re-interpretation of evidence from managed portfolios, *Review of Financial Studies* 6, 1-21.

Fama, Eugene, 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.

Fama, Eugene, and Kenneth French, 1992, The cross section of expected stock returns, *Journal of Finance* 47, 427-465.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene, and Kenneth French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153-193.

Fama, Eugene, and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Goetzmann, William, and Roger Ibbotson, 1994, Do winners repeat? Patterns in mutual fund performance, *Journal of Portfolio Management* 20, 9-18.

Goetzmann, William, Massimo Massa, K. Geert Rouwenhorst, 2000, Behavioral factors in mutual fund flows, Working Paper, Yale School of Management.

Grinblatt, Mark, and Sheridan Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 42, 1977-1984.

Gruber, Martin , 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance 51, 783-810.*

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of performance, 1974-88, *Journal of Finance* 48, 93-130.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winner and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.

Lakonishok Josef, Andrei Shleifer, Richard Thaler and Robert Vishny, 1991, Window dressing by pension fund managers, *AEA Papers and Proceedings, 227-231.*

Teo, Melvyn and Sung-Jun Woo, 2001, Style effects, Working Paper, Harvard University.

Warther, Vincent A., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209-235.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock picking talent, style, transactions costs and expenses, *Journal of Finance 55, 1655-1695.*

White, Halbert., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-838.

Zheng, Lu, 1999, Is money smart? A study of mutual fund investors' fund selection ability, *Journal of Finance* 54, 905-935.

7 Appendix

This section details the algorithm we use to estimate the equity style of a fund for the fund-years which have missing style information. The algorithm consists of the following sequence of steps in descending order of priority:[28]

1. Funds classified by Strategic Insight, Wiesenberger, and ICDI as sector funds, bond (including municipals) funds, money market funds, global funds, and international funds are omitted.

[28] For example, if a fund has "small cap index" in its name, then we label it a small-cap blend fund, even if its Strategic Insight code is SCG.

2. Funds classified by Strategic Insight as flexible funds (FLX), balanced funds (BAL), principal return funds (EPR), and corporate income mixed funds (IMX) are domestic hybrid funds and are omitted.

3. If a fund name makes explicit reference to the style of stocks the fund invests in, it is assumed to invest in that style, e.g., Consulting Grp Capital Markets Large Cap Value Equity fund is a large-cap value fund. Also, balanced funds are domestic hybrid funds and index funds are blends. Note that domestic hybrid funds are omitted.

4. Funds classified by Strategic Insight as mid-cap growth funds (GMC) are mid-cap growth funds.

5. Funds classified by Strategic Insight as small-cap growth funds (SCG) are small-cap growth funds.

6. Funds classified by Strategic Insight as growth and income funds (GRI) or income and growth funds (ING) are large value funds.

7. Funds classified by Strategic Insight as aggressive growth funds (AGG) are large growth funds.

8. Funds classified by Strategic Insight as growth funds (GRO) are large blend funds.

9. Funds classified by Wiesenberger as stability, income and growth funds (S-I-G) or income funds (I) are domestic hybrid funds and are omitted.

10. Funds classified by Wiesenberger as growth funds (G) are large growth funds.

11. Funds classified by Wiesenberger as income and growth (I-G) are large blend funds.