Conditional Benchmarks and Predictors of Mutual Fund Performance

Scott Cederburg* Michael S. O'Doherty† N. E. Savin‡ Ashish Tiwari§

March 1, 2018¶

Abstract

Recent studies link mutual fund performance to measures of active management, and this evidence often takes the form of large spreads in unconditional alphas for characteristic-sorted portfolios. Unconditional benchmarks can, however, produce misleading inferences on managerial skill for strategies that exhibit substantial turnover and unstable factor exposures. We propose a performance attribution model that accounts for predictable changes in portfolio style. Compared to existing methods, our benchmarks yield superior tracking performance and a more powerful statistical assessment of abnormal returns. We reevaluate six active management proxies using our method and conclude that these measures are largely unrelated to managerial ability.

JEL classifications: G11, G23

*Eller College of Management, University of Arizona. Email: cederburg@email.arizona.edu.

†Trulaske College of Business, University of Missouri. Email: odohertym@missouri.edu.

‡Tippie College of Business, University of Iowa. Email: gene-savin@uiowa.edu.

§Tippie College of Business, University of Iowa. Email: ashish-tiwari@uiowa.edu.

¶We thank Yinfei Chen, Brad Goldie, George Jiang, Haim Kassa, Jerchern Lin, Alexey Malakhov, Marno Verbeek, Xiaolu Wang, Ryan Whitby, and seminar participants at Iowa State University, Macquarie University, Miami University, the University of Arkansas, the University of Iowa, the University of New South Wales, Utah State University, the 2016 Financial Management Association meetings (Las Vegas), the 2016 India Finance Conference (Ahmedabad), and the 2017 Asset Management Conference (Berlin) for helpful comments and suggestions.

1 Introduction

Recent studies on the performance of equity mutual funds have produced an extensive list of apparently successful strategies for identifying superior funds. In particular, several fund-level characteristics that can be interpreted as managerial activeness proxies have been linked to mutual fund performance. These proposed predictors include industry concentration (Kacperczyk, Sialm, and Zheng (2005)), unobserved actions of funds (Kacperczyk, Sialm, and Zheng (2008)), active share (Cremers and Petajisto (2009)), mutual fund R^2 (Amihud and Goyenko (2013)), active weight (Doshi, Elkamhi, and Simutin (2015)), and fund return volatility (Jordan and Riley (2015)). Taken together, these studies imply that it is quite feasible to ex ante identify skilled fund managers using observable fund characteristics.[1]

In this paper, we propose a new method to evaluate the conditional performance of characteristic-sorted portfolios of mutual funds and apply this framework to the literature on predictors of fund performance. Our benchmarking procedure leads to dramatically different conclusions regarding the ex ante identification of superior managerial skill. For some background, a common empirical approach in the papers cited above is to sort mutual funds into portfolios based on the proposed predictor variable and evaluate the performance of the resulting trading strategies. This analysis typically involves computing style-adjusted returns via time-series regressions of the portfolio returns on market, size, value, and momentum factors (i.e., Carhart (1997) four-factor model regressions). Realized portfolio returns are effectively decomposed into a component that reflects the return that could be earned from exposures to the four passive benchmark factors and a residual component (i.e., alpha). The key evidence for distinguishing a useful predictor is that the extreme portfolios exhibit an economically large and statistically significant difference in alphas.[2]

A potential concern with the findings outlined above relates to the benchmarking procedure used to style-adjust the portfolio returns. Specifically, for each of the proposed predictors, the identity of the "skilled" funds changes through time. The associated strategies are designed to maintain exposure to the skilled managers by trading in and out of individual mutual funds over the sample period. As the portfolios are rebalanced each month, however, these strategies can exhibit

[1]Consistent with the practice in most prior studies on mutual funds, we use the term "skill" to refer to the net fund alpha, i.e., the average abnormal fund return net of expenses and fees. Recently, Berk and van Binsbergen (2015) have argued that a fund manager's skill is reflected in a value added measure, namely, the fund's gross return in excess of its benchmark multiplied by the assets under management. Because our paper specifically addresses prior studies that examine predictors of fund alphas, our focus is on the proper assessment of fund performance as reflected in the alpha measure.

[2]Kacperczyk and Seru (2007) show that a mutual fund manager's reliance on public information, which can also be interpreted as a proxy for activeness, has predictive content for mutual fund performance. The findings in this study, however, are not based on the unconditional Carhart (1997) regression approach.

pronounced shifts in both the identity of the constituent funds and their underlying factor exposures. Importantly, these changes in portfolio loadings can occur solely as a result of portfolio turnover, even if the underlying mutual funds are not adjusting their factor exposures over time. These style dynamics are not reflected in the standard Carhart (1997) four-factor benchmarking approach, which assigns constant exposures to a given strategy over the full sample period. Consequently, the reported alphas and apparent managerial skill may simply reflect portfolio style drift and a benchmark model that performs poorly in tracking the strategy returns.

To address this concern, we extend the conventional conditional performance evaluation framework to account for these predictable changes in style exposures. Our approach builds on the traditional implementation of conditional performance evaluation (e.g., Ferson and Schadt (1996), Ferson (2010), and Ferson (2013)), in which one or more of a portfolio's factor loadings is modeled as a linear function of state variables, such as the dividend yield, default spread, and term spread. These variables have a long history of use in forecasting asset returns (e.g., Fama and French (1989)), and the motivation for including them in a conditional benchmark is to control for mechanical factor timing strategies based on publicly available information. As Ferson and Schadt (1996) explain, any portfolio strategy that can be replicated using such information should not be deemed as having superior performance. Our innovation is to replace or complement the traditional conditioning variables with another set of instruments based on lagged factor loading estimates for a strategy's constituent mutual funds. We specifically propose using the portfolio-weighted average lagged factor loadings across funds held in a particular portfolio as instruments for that portfolio's exposures in a Carhart (1997) model regression. The use of lagged loadings as instruments for conditional factor exposures was first proposed by Boguth, Carlson, Fisher, and Simutin (2011), who employ this method to reexamine the performance of stock momentum strategies.

Our proposed approach has several attractive features. First, it is easy to implement, as it is based on standard time-series regression methods used in the mutual fund literature. Second, the lagged factor loadings allow the researcher to incorporate a powerful source of publicly available information in predicting future style exposures and benchmarking performance. In particular, these instruments provide a simple way to pick up high-frequency shifts in style (e.g., at portfolio rebalancing dates) that would be missed by traditional conditioning variables, such as the dividend yield, which tend to be much more persistent in nature. In our empirical applications, we show that models with lagged loadings as instruments exhibit pronounced improvements in tracking strategy returns over unconditional Carhart (1997) model regressions and models with traditional conditioning variables. Third, our method leads to an intuitive decomposition of a given strategy's unconditional alpha, which tends to be the focus of prior literature as noted above, into performance in security selection (i.e., the conditional alpha), factor timing, and volatility timing. As noted by Ferson and Mo (2016), measures of selectivity that ignore managerial ability in timing either factor returns or factor volatility suffer from an omitted variable bias. Boguth, Carlson, Fisher, and Simutin (2011) make a similar point regarding the importance of market timing and volatility timing in returns-based instrumental variables tests. Finally, because our conditional models with lagged factor loading instruments lead to improved tracking performance (i.e., markedly higher time-series regression R^2s), this approach also produces extremely precise conditional alpha estimates. As such, our tests have increased statistical power to identify skill in security selection among the strategies of interest.

To demonstrate the usefulness of our approach, we reevaluate an important set of existing results on the predictive content of mutual fund manager activeness. The general conclusion from existing studies is that fund managers who take more active bets in their portfolios tend to outperform. For example, Cremers and Petajisto (2009) and Doshi, Elkamhi, and Simutin (2015) develop holdings-based measures of a manager's tendency to deviate from benchmark portfolio weights and find that active share and active weight, respectively, are positive predictors of fund alphas. Similarly, Kacperczyk, Sialm, and Zheng (2005) show that funds that are more concentrated across industries earn higher alphas. Amihud and Goyenko (2013) produce complementary results using a measure of activeness based on mutual fund returns. They specifically find that funds with low R^2 values from Carhart (1997) four-factor regressions subsequently outperform funds with high R^2s. Finally, Kacperczyk, Sialm, and Zheng (2008) introduce the return gap measure, which can be interpreted as a proxy for activeness, and show that funds earning higher returns than a passive investment in their recently disclosed holdings also earn higher future alphas.[3] One recent paper producing seemingly contradictory results is Jordan and Riley (2015), who find that lagged mutual fund volatility exhibits a pronounced inverse relation to unconditional alpha.

This literature provides a natural arena to apply our conditional performance evaluation methods, given the prior emphasis on unconditional Carhart (1997) model regressions. As such, we examine the ability of R^2, active share, active weight, volatility, industry concentration, and return gap to predict mutual fund performance. Our empirical results focus on evaluating monthly rebalanced decile portfolios formed on each of these six predictor variables. Following the convention in prior studies, much of our empirical analysis centers on assessing the performance of

[3]The positive association between activeness and future performance is not isolated to mutual funds. Titman and Tiu (2011) and Sun, Wang, and Zheng (2012) show that hedge funds following unique investment strategies have superior investment ability.

hypothetical strategies that take long and short positions in the extreme portfolios sorted on a given characteristic.

For our primary tests, we extend the sample period from each of the original studies through December 2015. The point estimates of unconditional alpha across the six long-short activeness strategies are economically large, ranging from 1.37% to 6.17% per year. All six of these estimates are statistically significant at conventional levels. We further demonstrate, however, that the extreme portfolios sorted on the proxies for mutual fund activeness exhibit considerable portfolio turnover, which leads to predictable shifts in style exposures over the sample period. As an example, the low-R^2 and high-R^2 deciles in the R^2 strategy require annualized portfolio turnover of 135% and 147%, respectively. We also find that the superior performance for each of the six long-short strategies is concentrated over short sample windows, and these periods often correspond with discrete changes in investment styles due to portfolio rebalancing. These results raise the possibility that the prior evidence on superior performance for the strategies of interest is attributable to poorly specified benchmark models, rather than skill in security selection.

Consistent with this possibility, we find limited evidence that fund activeness is associated with future performance relative to our conditional benchmarks that account for predictable shifts in style. The conditional alphas for the long-short portfolios range from 0.07% to 1.61% per year, with an average reduction in magnitude of 61% in comparison to their corresponding unconditional estimates. Across the six predictors, the only statistically significant conditional alphas are earned by the active weight (10% significance level) and return gap (1% significance level) strategies. We further demonstrate that our findings are robust to alternative approaches to constructing instruments for our conditional models. Across a wide range of empirical specifications, conditional models deliver superior tracking performance relative to unconditional benchmarks and reliably lead to significant reductions in abnormal returns for the long-short activeness portfolios. As such, the evidence suggests that proxies for managerial activeness bear little relation to security selection ability, in contrast to the claims in prior studies.

We also compare our conditional models based on lagged factor loading instruments to the traditional approach to conditional performance evaluation, in which portfolio factor loadings are modeled as a function of macroeconomic variables. We find that the models featuring lagged loadings exhibit meaningful advantages in terms of model fit, as evidenced by their adjusted-R^2 values. In many cases, the conditional alpha estimates differ considerably across the two approaches, suggesting that our conclusions on managerial ability would be missed by simply following the conditional approach adopted in prior literature.

Finally, we present decompositions of the unconditional alphas for the six long-short strategies into factor timing, volatility timing, and security selection effects. In most cases, we find that the large unconditional alphas for these portfolios are primarily attributable to factor timing. That is, whereas the strategy portfolios tend to have relatively high conditional exposures to the benchmark factors in periods when these factors earn high returns, there is limited evidence that their superior performance is linked to ability in security selection. An important question is whether this apparent success in factor timing is attributable to managers of the underlying funds skillfully shifting their exposures over time or simply to a given strategy's rebalancing procedure. Our results are more consistent with the latter explanation. For those strategies exhibiting significant factor timing ability, most of this performance is concentrated over a few sample months in the period just preceding and/or subsequent to the crash in technology stocks in March 2000. Using a detailed analysis of portfolio holdings, we show that, while several of the mutual fund strategies successfully adjust their weights in technology stocks over this period, the constituent mutual funds do not. Thus, none of the individual predictors appears to robustly identify managers with factor timing ability.

The paper contributes to an extensive literature on the performance evaluation of managed portfolios. From an economic perspective, we provide evidence that the link between managerial activeness and mutual fund performance is tenuous at best. The nature of our results indicates that properly specifying conditional benchmarks is critical for making inferences about skill in active management. Conceptually, our conditional benchmarking approach is similar in spirit to those of Daniel, Grinblatt, Titman, and Wermers (1997) and Ferson and Mo (2016), which use fund holdings information to track the dynamic investment styles of the funds being evaluated.[4] However, unlike their benchmarks, our conditional models evaluate performance based on fund returns, which can be advantageous in accounting for non-equity positions, stale holdings data, and realized trading costs. Ferson and Mo (2016) also decompose fund performance into security selection, factor timing, and volatility timing components using holdings data, and our approach complements theirs by providing a returns-based alternative for this decomposition. Finally, our study builds on the conditional performance evaluation framework introduced by Ferson and Schadt (1996), and we find that lagged factor loading instruments considerably outperform traditional

[4]In related work, a number of studies have sought to model the dynamic factor exposures of mutual funds and hedge funds using a variety of techniques. Mamaysky, Spiegel, and Zhang (2008) use a Kalman filter-based model to track the dynamic factor loadings of mutual funds. Bollen and Whaley (2009) employ an optimal changepoint detection framework to identify potential shifts in the factor exposures of hedge funds. More recently, Patton and Ramadorai (2013) use high frequency conditioning variables to capture within-month variation in the risk exposures of hedge funds.

instruments in tracking portfolio returns. A byproduct of this improved tracking ability is an increase in test power, such that this method is an important tool for investors and researchers who wish to identify significant predictors of managerial skill.

The remainder of the paper is organized as follows. Section 2 details our sample selection procedures and construction of the mutual fund strategies based on proxies for managerial activeness. Section 3 introduces our estimation approach for performance evaluation and contrasts it with existing methods. Section 4 presents our results on the performance of the mutual fund portfolios, and Section 5 concludes.

2 Data

Section 2.1 provides details on our mutual fund sample. Section 2.2 introduces the fund-level proxies for active management and describes the construction of mutual fund portfolios formed on these activeness measures.

2.1 Sample construction

We obtain data on monthly mutual fund returns from the CRSP Survivor-Bias-Free US Mutual Fund Database for the period April 1980-December 2015. These returns are net of fees, expenses, and brokerage commissions but before any front-end or back-end loads. We convert all net returns to excess returns by subtracting the corresponding risk-free rate.[5] We also collect data on fund characteristics, including total net assets, expense ratio, turnover, and percentage of the portfolio invested in common stocks, preferred stocks, bonds, cash, and other securities. We use the MFLINKS database to identify funds with multiple share classes and combine these share classes into portfolios. A fund's total net assets for a given period is the sum of total net assets across share classes, and the fund's returns and other characteristics are asset-weighted averages.

To limit the sample to domestic actively managed equity funds, we follow the approach in Doshi, Elkamhi, and Simutin (2015) and screen on the investment style codes from CRSP (i.e., crsp_obj_cd). We check the dataset for index funds and eliminate these observations from the sample. We also eliminate balanced funds, bond funds, international funds, sector funds, funds with missing names, and funds that have less than 80% of their holdings on average in common

[5] We obtain data on daily and monthly factor returns (i.e., the market, size, value, and momentum factors from the Carhart (1997) four-factor model) and the risk-free rate from Kenneth French's website. See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. We thank Kenneth French for making these data available.

stocks. We include two additional screens to address potential concerns related to incubation bias (e.g., Evans (2010)). First, we delete any fund-month observation that precedes the fund's first offer date as recorded in the CRSP database. Second, we only include funds in the strategy portfolios if their total net assets exceeds \$15 million at the portfolio formation date.

2.2 Proxies for active management

In our empirical tests, we focus on six widely used measures of active management. Each of these variables has been shown in prior literature to predict cross-sectional differences in fund performance. Several of the activeness measures require data on fund holdings, which we obtain from the Thomson Reuters Mutual Fund Holdings Database and link to the CRSP mutual fund database using MFLINKS. The fund holdings database also contains stock-level identifiers, which we use to link to the CRSP Monthly Stock File to obtain SIC codes, exchange codes, market capitalizations, and returns for individual holdings.

We construct each of the following fund-level proxies for activeness on a monthly basis. The start of the sample period for a given measure corresponds to the start date of the portfolio analysis in the original article introducing that characteristic as a predictor of mutual fund performance.

- $\boldsymbol{R^2}$. Following Amihud and Goyenko (2013), we estimate the R^2 from a regression of mutual fund excess returns on the four Carhart (1997) factors using the prior 24 months of data. We require funds to have a valid return for each of the prior 24 months, and the sample period for R^2 is January 1990-December 2015.

- **Active share (*AS*)**. Following Cremers and Petajisto (2009), active share for mutual fund i at the end of month t is defined as

$$AS_{i,t} = \frac{1}{2}\sum_{j=1}^{J} |w_{i,t}^{j} - w_{i,t}^{j,b}|, \tag{1}$$

where $w_{i,t}^{j}$ is the equity portfolio weight of stock j in fund i, $w_{i,t}^{j,b}$ is the equity portfolio weight of stock j in the fund's benchmark index, and the sum is taken across the universe of all stocks. Cremers and Petajisto (2009) consider a total of 19 candidate benchmark indexes, and a given fund's benchmark is defined as the one that minimizes its active share. We use an updated version of active share from Martijn Cremers' website that allows for 58 potential benchmark indexes.[6] The sample period for active share is January 1990-December 2015.[7]

- **Active weight ($\boldsymbol{AW}$).** Following Doshi, Elkamhi, and Simutin (2015), we compute active weight for mutual fund i at the end of month t as

$$AW_{i,t} = \frac{1}{2}\sum_{j=1}^{J} |w_{i,t}^{j} - w_{i,t}^{j,m}|, \tag{2}$$

where $w_{i,t}^{j,m}$ is the equity portfolio weight of stock j in a market capitalization-weighted portfolio of the stocks in fund i. We require funds to hold at least 10 stocks, and the sample period for active weight is April 1980-December 2015.

- **Volatility ($\boldsymbol{Vol}$).** Following Jordan and Riley (2015), we compute volatility as the standard deviation of daily mutual fund returns over the prior 12 months. We require funds to have a valid return for each trading day over that period. Data on daily fund returns are from the CRSP daily mutual fund return file, which starts in September 1998. We use January 1999-December 1999 as the initial estimation window for computing fund-level volatility, and the sample period for this measure is January 2000-December 2015.

- **Industry concentration ($\boldsymbol{ICI}$).** Following Kacperczyk, Sialm, and Zheng (2005), we compute industry concentration index for mutual fund i at the end of month t as

$$ICI_{i,t} = \sum_{n=1}^{10} (w_{i,t}^{n} - \bar{w}_{t}^{n})^2, \tag{3}$$

where $w_{i,t}^{n}$ is equity portfolio weight of industry n in fund i and $\bar{w}_{t}^{n}$ is the weight in industry n for the total stock market. We use data on the 48-industry classification, available on Kenneth French's website, and Table AI in the Appendix of Kacperczyk, Sialm, and Zheng (2005) to map stock-level SIC codes into the ten industries required to compute industry concentration. We require funds to hold at least 10 stocks, and the sample period for industry concentration is January 1984-December 2015.

- **Return gap ($\boldsymbol{RetGap}$).** Following Kacperczyk, Sialm, and Zheng (2008), we compute

[6] See http://activeshare.nd.edu. Cremers and Pareek (2016) present a detailed description of the data construction, and we thank Martijn Cremers for making these data available.

[7] Our results are robust to using Cremers and Petajisto's (2009) original measure of active share based on 19 benchmarks. These data are available on Antti Petajisto's website at http://www.petajisto.net/data.html. We focus on the updated version of active share because the original measure is only available through December 2009. Our results are also robust to using a version of active share based on each fund's self-reported benchmark, as in Petajisto (2013).

return gap as the difference between a fund's monthly gross return and the return of its most recently reported holdings, averaged over the prior 12 months. The Thomson Reuters Mutual Fund Holdings Database includes only common stock positions, and we adjust the fund holding returns to account for returns on other asset classes. We specifically assume that bonds and preferred stocks earn the return of the Barclays U.S. Aggregate Bond Index and that cash and other assets earn the risk-free rate. We require funds to hold at least 10 stocks, and the sample period for return gap is January 1984-December 2015.

Our empirical results are based on six sets of test portfolios. For each of the proxies for managerial activeness, we construct a standard set of decile portfolios based on a one-way sort. The portfolios are equal weighted and rebalanced monthly. As discussed below, our conditional benchmarking approach uses data on fund holdings to construct lagged factor loading instruments. We therefore only include funds in the portfolios for a given month if they have reported their holdings within the prior 12 months. In the tables that follow, we often focus on the performance of the top and bottom deciles for each measure of activeness. Following the convention in the literature, we also assess the performance of hypothetical long-short strategies that take positions in each of these extreme groups. Amihud and Goyenko (2013) and Jordan and Riley (2015) show that R^2 and volatility, respectively, are negatively related to future fund performance. As such, we form low-minus-high strategies based on these two predictors. The remaining four proxies are positive predictors of performance, so we consider a high-minus-low portfolio based on each of these variables.

3 Performance evaluation methods and conditioning variables

In this section, we introduce our method for conditional performance evaluation and contrast it with the unconditional and conditional factor regression approaches traditionally applied in the literature. Section 3.1 develops our conditional approach to assessing the performance of mutual fund strategies. Section 3.2 discusses conditioning variables based on lagged estimates of strategy factor loadings, and Section 3.3 provides information about the traditional instruments that are used in some of our conditional models.

3.1 Conditional approach to assessing performance

The returns for mutual fund strategies are often benchmark-adjusted using an unconditional Carhart (1997) four-factor model regression,

$$R_{i,t} = \alpha_i^U + \beta_i^U R_{MKT,t} + s_i^U R_{SMB,t} + h_i^U R_{HML,t} + u_i^U R_{UMD,t} + \varepsilon_{i,t}, \tag{4}$$

where $R_{i,t}$ is the excess return for portfolio i in month t, $R_{MKT,t}$ is the return on the market factor, $R_{SMB,t}$ and $R_{HML,t}$ are the size and value factors of the Fama and French (1993) three-factor model, and $R_{UMD,t}$ is a momentum factor. In the spirit of Sharpe (1992), this approach serves to decompose the realized return for a given strategy into two components—a component that reflects the return that could be earned from suitable exposures to the four passive benchmark factors and a residual component (i.e., the unconditional alpha, α_i^U) that is often interpreted as managerial skill. Such an approach, however, inherently assumes that the appropriate benchmark portfolio retains constant exposures to the underlying factors. Whereas this assumption is arguably reasonable for benchmarking an individual mutual fund, the results in the literature for predictors of mutual fund performance are typically based on portfolios of funds. These portfolios require trading in and out of individual funds, which often leads to considerable variation in factor exposures.

Ideally, the benchmarks applied to these strategies in assessing managerial performance would account for any predictable changes in factor exposures over time. Conditional benchmarks are likely to track a given strategy's returns better than unconditional benchmarks if factor exposures are time varying.[8] Further, applying the framework of Hansen and Richard (1987) to mutual fund performance evaluation, it is possible for fund managers to appear to have skill relative to unconditional benchmarks but to show no skill after conditioning on the investor information set. More specifically, we know from the asset pricing literature that an unconditional portfolio alpha may be a biased estimate of the conditional alpha if factor loadings vary systematically with the expected returns (i.e., "factor timing") or volatilities (i.e., "volatility timing") of the factors (see, e.g., Grant (1977), Jagannathan and Wang (1996), Lewellen and Nagel (2006), and Boguth, Carlson, Fisher, and Simutin (2011)).

In the context of mutual fund performance evaluation, the conditional alpha is a direct measure of skill in security selection. The unconditional alpha, on the other hand, will also reflect factor

[8] For a simple example, consider a strategy that invests entirely in mutual funds implementing value strategies over the first half of the sample period and invests entirely in growth-oriented funds over the second half of the sample. The resulting estimate of the value factor loading, $\hat{h}_i^U$, from the unconditional benchmark in equation (4) might be close to zero, but such a benchmark would likely perform poorly in tracking the strategy's returns.

timing and volatility timing effects. Factor timing has been extensively considered in the mutual fund literature, originating with Treynor and Mazuy (1966), Grant (1977), and Henriksson and Merton (1981).[9] In addition, Busse (1999) considers the ability of mutual fund managers to time market volatility. Success in factor timing and volatility timing adds value for investors, but timing effects for a given strategy portfolio may or may not indicate managerial timing skill.[10] We discuss this issue in detail in Section 4.5.

A standard conditional approach to assessing performance is to estimate a version of the Carhart (1997) model that allows factor loadings to vary over time. We assume that the conditional alpha is constant and that the conditional portfolio factor loadings are linear in a set of conditioning variables (e.g., $\beta^C_{i,t} \equiv \lambda_{i,0} + \lambda'_{i,1} Z^{MKT}_{i,t-1}$). Specifically, we measure the conditional alpha of each portfolio using the regression,

$$\begin{aligned} R_{i,t} = \alpha^C_i &+ (\lambda_{i,0} + \lambda'_{i,1} Z^{MKT}_{i,t-1}) R_{MKT,t} + (\gamma_{i,0} + \gamma'_{i,1} Z^{SMB}_{i,t-1}) R_{SMB,t} \\ &+ (\eta_{i,0} + \eta'_{i,1} Z^{HML}_{i,t-1}) R_{HML,t} + (\nu_{i,0} + \nu'_{i,1} Z^{UMD}_{i,t-1}) R_{UMD,t} + \varepsilon_{i,t}, \end{aligned} \tag{5}$$

where $Z^k_{i,t-1}$ is an $n_{i,k} \times 1$ vector of instruments that can vary across portfolios and factors. Importantly, $Z^k_{i,t-1}$ is in the investor information set at the beginning of period t.

This method of measuring mutual fund performance using conditional factor models was first developed by Ferson and Schadt (1996). The traditional approach in the literature is to use macroeconomic state variables, such as the dividend yield and interest-rate related variables, as instruments for portfolio factor loadings. Our innovation is to borrow from recent advances in the asset pricing literature and add lagged averages of estimated factor loadings for constituent funds as conditioning variables.

For some background on this approach, Lewellen and Nagel (2006) advocate measuring conditional portfolio risk with contemporaneous short-window regression betas to avoid problems associated with conditioning on a subset of the information available to investors. Boguth, Carlson, Fisher, and Simutin (2011) show that this method may cause an "overconditioning" bias because the short-window betas are not known to investors at the beginning of the period. Lagged portfolio

[9] A partial list of additional studies on market timing by mutual funds includes Henriksson (1984), Jagannathan and Korajczyk (1986), Ferson and Schadt (1996), Becker, Ferson, Myers, and Schill (1999), and Jiang, Yao, and Yu (2007). Avramov and Chordia (2006) account for predictability in the security selection and factor timing skills of managers.

[10] For a mutual fund that strategically shifts its exposures to the benchmark factors over time, the intercept from equation (4) will reflect ability in both timing and security selection. For a portfolio of mutual funds, however, an unconditional alpha that results from predictable changes in factor loadings attributable to portfolio rebalancing is not direct evidence of managerial skill in timing the factors.

loadings are, however, in the investor information set, and Boguth, Carlson, Fisher, and Simutin (2011) demonstrate that these variables serve as good instruments for factor exposures of stock portfolios in tests similar to equation (5).[11]

Estimating conditional benchmarks using lagged loadings as instruments has not yet been applied in the mutual fund literature, but the method is particularly well suited for this setting. Directly using recent factor loadings to predict exposures will perform better when the factor loading estimates have low levels of measurement error. Because mutual funds are diversified portfolios, the factor loadings of funds and strategies that form portfolios of funds can be estimated relatively precisely over short periods.[12] Further, the mutual fund strategies evaluated in the literature often require frequent rebalancing and changes in the identity of constituent funds. The lagged loading instruments we design in Section 3.2 are based on the estimated exposures of mutual funds that are currently in a given strategy portfolio, so these instruments rapidly adjust to the inclusion of new funds. Traditional instruments based on macroeconomic variables, on the other hand, tend to move at business cycle or lower frequencies, such that they may provide a poor fit to the short-term movements in factor loadings that result from rebalancing. Finally, choosing instruments for conditional models is subjective and can lead to data-mining concerns (e.g., Ferson, Sarkissian, and Simin (2008) and Cooper and Gubellini (2011)), whereas using lagged loading estimates as conditioning information for factor loadings removes much of the subjectivity from the method.

In our empirical analysis, we measure the unconditional and conditional performance of portfolios formed on proxies for mutual fund manager activeness. We estimate the models in equations (4) and (5) using the generalized method of moments (GMM). Each of these regression models is exactly identified, and the GMM parameter estimates correspond to ordinary least squares estimates. We estimate standard errors using the approach in White (1980) to account for potential heteroskedasticity. For each measure of activeness, we consider a hypothetical long-short strategy that is predicted to produce positive unconditional performance based on results in prior literature. Our main tests assess whether the conditional alpha of the long-short portfolio, α_{Δ}^{C}, for each sorting variable is equal to zero. We also assess whether these conditional alphas are significantly smaller than their unconditional counterparts by testing the null hypothesis $\alpha_{\Delta}^{C} = \alpha_{\Delta}^{U}$ against the

[11] An alternative approach that employs lagged portfolio betas as direct proxies, rather than as instruments for conditional portfolio risk exposures, is also problematic if portfolio risk changes predictably between the prior period and the current holding period (see, e.g., Chan (1988), Grundy and Martin (2001), and Boguth, Carlson, Fisher, and Simutin (2011)).

[12] One potential concern is that stale pricing can lead to downward biased estimates of factor loadings (e.g., Dimson (1979)), particularly for portfolios of small-cap funds. The regression model in equation (5) implicitly controls for this bias, as lagged loadings are used as instruments, which are rescaled in estimating the model, rather than as direct proxies for conditional exposures.

alternative $\alpha_{\Delta}^{C} < \alpha_{\Delta}^{U}$.[13] Given that investors cannot take short positions in mutual funds, we also perform analogous tests for the performance of the long leg portfolio for each strategy. Finally, we compare inferences from conditional models that use lagged factor loading instruments with models using traditional instruments from prior literature in equation (5).

3.2 Lagged factor loading instruments

Our base specification uses lagged three-month factor loading estimates as conditioning variables in equation (5). To construct these lagged factor loading instruments for each strategy portfolio, we first estimate the loadings for each of the mutual funds in the portfolio. We develop two methods for estimating fund factor loadings. The first approach directly uses daily fund-level excess returns and factor returns to estimate an unconditional Carhart (1997) model regression over the most recent three-month period. The CRSP daily mutual fund return file starts in September 1998. Thus, these factor loading estimates are available for the full sample period for the volatility strategy, but they are unavailable in the early part of the sample for the remaining five strategies. To overcome this data limitation, the second approach estimates fund factor loadings using data on fund holdings and returns from the CRSP Daily Stock File. Specifically, we first estimate unconditional Carhart (1997) factor loadings for each stock over the prior three-month period. We then use the most recently reported holdings for each mutual fund to estimate fund factor loadings as the weighted average of the estimated stock-level factor loadings. In computing these fund-level loadings at the end of month $t-1$, we adjust the stock-level weights to account for changes in the market value of securities between the holdings date and time $t-1$. Given the availability of data on mutual fund holdings and daily stock returns, we are able to produce these fund factor loading estimates throughout the full sample for each strategy. Finally, the average factor loadings across funds in a strategy portfolio are estimates of the lagged three-month loadings β_{t-1}^{L3}, s_{t-1}^{L3}, h_{t-1}^{L3}, and u_{t-1}^{L3} for the portfolio.[14]

Each set of lagged factor loading instruments has potential advantages and drawbacks relative to its alternative. A clear advantage of the holdings-based measure is that it is available for each strategy's full sample period, whereas the instruments based on daily mutual fund returns data are first available in January 1999. Within a given sample period, either approach could produce

[13] The primary advantage of our GMM estimation approach is that we are able to easily conduct this cross-equation hypothesis test. The test specifically involves estimating the models in equations (4) and (5) in a single GMM procedure. See Appendix A.5 in Boguth, Carlson, Fisher, and Simutin (2011) for details.

[14] The approach to calculating a portfolio's betas based on the estimated betas of its constituent assets is referred to as the "lagged component" approach by Boguth, Carlson, Fisher, and Simutin (2011).

better predictors of realized strategy factor loadings. If fund holdings data provide a more up-to-date view of the current portfolio relative to the past three months of historical fund returns data, the holdings-based instruments may outperform the measures based on fund returns. Alternatively, the method that directly uses fund returns may better forecast strategy loadings if holdings are relatively stale or if funds rely on dynamic trading strategies that affect fund betas but are not reflected in a snapshot of fund holdings. Further, holdings data only cover stock positions, whereas factor loading estimates based on fund returns reflect holdings of cash and other non-stock assets.

Figure 1 shows lagged factor loading instruments for the low-R^2 strategy portfolio. The panels plot the holdings-based and returns-based instruments for each of the Carhart (1997) four factors. As discussed above, the returns-based instruments are available beginning in January 1999, and the holdings-based measures are shown for the January 1990-December 2015 period. The two sets of instruments are visually quite similar across the factors. The most noticeable difference is that the lagged factor loadings from the returns-based approach tend to be shifted toward zero relative to the holdings-based measures. This finding primarily reflects fund holdings in cash, which reduces the magnitude of fund factor exposures. The conditional performance evaluation method in equation (5) implicitly adjusts for this effect by rescaling the lagged factor loading instruments to model the factor exposures of a given strategy portfolio. The two instruments in each panel otherwise display very similar patterns for long-term and short-term shifts in factor exposures. Figure 1 also provides preliminary evidence that accounting for time variation in strategy factor exposures is important given the volatility of the factor loading estimates over time.

We further study the efficacy of returns-based and holdings-based instruments in Section 4.3 by comparing the fit of conditional models that use a given set of instruments. Returns-based instruments tend to provide better model fit over the January 1999-December 2015 period, but both approaches yield significantly better tracking performance relative to the unconditional models. In our base specification, we use returns-based instruments for the volatility strategy because they are available for the full sample period, and we use holdings-based instruments for the remaining strategies. We also investigate shorter-term and longer-term instruments that are formed analogously to the three-month instruments, but have alternative measurement periods. In particular, we construct lagged instruments using daily return regressions over periods ranging in length from one month to 12 months, and we form 24-month factor loading instruments for each portfolio using monthly return regressions. In some empirical tests, we include both shorter-term and longer-term lagged factor loadings as conditioning variables for the matching factor loading in equation (5).

3.3 Traditional instruments

In Section 4.4, we contrast conditional models that are based on lagged factor loading instruments with models that employ standard conditioning variables. These traditional state variables include the dividend yield, default spread, and term spread. The dividend yield is the sum of dividends accruing to the CRSP value-weighted market portfolio over the prior 12 months divided by the current index level. The default spread is the difference in yields between Moody's Baa- and Aaa-rated bonds, and the term spread is the difference between the 10-year Treasury constant maturity rate and the one-year Treasury constant maturity rate. All bond yields are obtained from the Federal Reserve Bank of St. Louis website.[15]

4 Results

In this section, we apply the empirical methods introduced in Section 3 to examine the performance of the decile portfolios formed on the activeness proxies. Section 4.1 provides unconditional performance measures and discusses strategy portfolio characteristics that motivate the use of conditional models to evaluate portfolio performance. Section 4.2 presents our base conditional performance evaluation results, and Section 4.3 investigates conditional models with alternative lagged factor loading instruments. Section 4.4 compares our approach based on lagged factor loadings to those adopted in prior studies using traditional instruments. Section 4.5 introduces a decomposition to understand the sources of unconditional alphas for the mutual fund portfolios.

4.1 Unconditional performance evaluation

Table I reports unconditional Carhart (1997) model regression estimates for the long-short strategies of interest. The unconditional alpha for each portfolio is reported in percentage per year. Panel A of Table I shows results with sample periods that match the original study for each measure, and Panel B extends each of these samples through December 2015. Panel A produces evidence that R^2, active weight, fund return volatility, industry concentration index, and return gap are significant predictors of fund performance, whereas the unconditional alpha estimate for the active share strategy is positive but statistically insignificant at conventional levels. The magnitude and statistical significance of these results are in line with the findings in the original studies.[16]

[15] See http://research.stlouisfed.org/fred2/.

[16] Given that we use consistent sample formation screens and portfolio formation methods across the measures, our results may not perfectly replicate those in prior studies. The alpha estimates reported in Panel A of Table I are, nonetheless, close in economic magnitude to the corresponding figures from prior literature. Our tests also match the

Panel B of Table I presents estimates of unconditional alphas and factor loadings over the longer sample periods that are the focus of our main analysis. The results suggest that extending the sample periods of the original studies does little to degrade strategy performance or change our statistical assessment of the unconditional alpha estimates (with the exception of the active share strategy). For example, the hypothetical long-short portfolios based on volatility and return gap earn abnormal returns of 6.17% and 1.76% per year, respectively. Both estimates are statistically significant at the 1% level using a two-tailed test. Further, the unconditional alpha estimates of the R^2, active weight, and industry concentration strategies are statistically significant at the 5% level, and the active share portfolio has significant positive performance at the 10% level. The economic magnitudes of these six abnormal return spreads are large, ranging from 1.37% to 6.17% per year across the measures.

To provide a deeper understanding of the strategies of interest, Table II presents portfolio characteristics. Panel A shows average net return, average gross return, and standard deviation of net return for the extreme strategy portfolios. Consistent with the results for unconditional alphas from Panel B of Table I, the activeness measures produce substantial spreads in average fund returns. For example, the low-minus-high R^2 portfolio provides an average net return of 2.54% per year and an average gross return of 2.90%. Panel B reports average characteristics of the mutual funds held in each portfolio. Average expense ratios tend to be somewhat larger for categories that indicate higher managerial activeness, but the magnitudes of these differences are small in comparison to the average return differences. Fund-level turnover is also generally higher for strategy portfolios with more active managers according to the measures.

More importantly, Panel C of Table II shows that investments based on the activeness measures require considerable strategy-level turnover. For example, an investor pursuing Amihud and Goyenko's (2013) proposed strategy of investing in low-R^2 mutual funds would see turnover of 135% per year. The high-R^2 portfolio exhibits even higher annualized turnover at 147%. These results highlight the dynamic nature of these strategies and have potentially critical implications for the evaluation of their performance. In particular, the high turnover among the extreme R^2 groups suggests that the identity and characteristics of the constituent funds are likely to change quite significantly over time. As such, the unconditional risk exposures presented in Table I used to benchmark strategy performance may mask considerable time variation in style exposures for the R^2 portfolios.

original studies in terms of statistical inference. Notably, our results using active share decile portfolios are consistent with Cremers and Petajisto's (2009) unconditional Carhart (1997) alpha for a high-minus-low active share quintile strategy that is positive, but statistically insignificant (see, e.g., Table 8 of their paper).

We see direct evidence of this effect in Figure 1 and Panel D of Table II. Figure 1 plots the three-month lagged loading estimates for each of the four factors in the Carhart (1997) model for the low-R^2 portfolio. Several of the portfolio factor loadings show pronounced shifts and trends across time, which provides direct motivation for using a conditional version of the Carhart (1997) model for performance evaluation. For example, the lagged three-month, holdings-based loading on the value factor ranges from -0.54, indicating a strong growth tilt, to 0.64, suggesting a pronounced exposure to value stocks. The loadings for the market, size, and momentum factors also show shifts that are economically large in magnitude. Note that these large swings in loadings are unlikely to be attributable to estimation error. Factor loadings for mutual funds tend to be estimated quite precisely compared to, say, individual stocks, and the estimates presented in Figure 1 are also averages across mutual funds in a given portfolio.[17]

Panels C and D of Table II show that these issues are relevant for the portfolios formed on alternative predictors of performance. Although several of the strategies require less trading in comparison to the R^2 portfolios, the annualized turnovers for the extreme decile portfolios are still substantial and exceed 75% with the exception of the active share strategy. The standard deviations of lagged portfolio factor loadings shown in Panel D are generally similar across the measures. There is some evidence of asymmetry in the reported volatilities, as the active decile for each proxy tends to exhibit greater variance in its style exposures. We demonstrate below that the alphas for these more active deciles are also more affected by the conditional tests.

To gain additional perspective on the predictive ability of the variables, we plot the time series of returns for the corresponding long-short strategy portfolios in Figure 2. The performance of several of the strategies is highly concentrated in the years 1999 and 2000, a period well known for the pronounced run-up and subsequent crash in the prices of technology stocks. In particular, the three highest monthly returns for the active weight, industry concentration index, and return gap measures are realized in a four-month span from November 1999 to February 2000, and these months also account for some of the highest returns for the R^2 and active share strategies. We also see instances of extreme negative performance around this period, with the low-minus-high volatility portfolio earning returns of -26.3% in February 2000 and -17.6% in June 2000. Importantly, the period over which the mutual fund strategies realize this extreme performance is marked by volatile

[17]For the low-R^2 strategy shown in Figure 1, we compute the average fund-level standard errors each month for the lagged three-month, returns-based loadings on the market, size, value, and momentum factors over the period January 1999-December 2015. The time-series averages of these values are 0.10, 0.14, 0.20, and 0.13, respectively. The average standard errors for the instruments shown in Figure 1, accounting for error correlation across funds, are 0.02, 0.04, 0.05, and 0.03. Across the six sets of portfolios considered in the paper, we find that the standard errors of the lagged loadings are typically higher for the more active deciles.

factor returns, such that properly measuring conditional factor exposures is critical for assessing managerial skill.

4.2 Conditional performance evaluation

Table III reports results from measuring strategy performance using the conditional performance evaluation approach. Each panel shows parameter estimates corresponding to one of the six individual predictor variables. The unconditional model in each case represents the specification in which factor loadings are constant. These regressions correspond to the unconditional Carhart (1997) model results in Panel B of Table I. The conditional models in each panel use the three-month lagged factor loading instruments to capture time variation in factor exposures. We report alphas and adjusted R^2s for the extreme portfolios and present parameter estimates for the factor exposures in the conditional models. We also show the unconditional and conditional alphas for the long-short strategy portfolio. Finally, for each of the long (i.e., the decile expected to outperform based on prior literature) and long-short portfolios, we report a p-value from a one-tailed test of the null hypothesis that the conditional alpha is equal to the corresponding unconditional alpha against the alternative that the conditional alpha is less than the unconditional alpha.

We begin with a detailed analysis of the R^2 strategy results in Panel A of Table III. The annualized unconditional Carhart (1997) model alpha of this strategy over the 1990-2015 period is 1.91% with a standard error of 0.84%. The conditional model introduces the three-month lagged factor loading instruments for the low-R^2 and high-R^2 portfolio loadings. Each of the eight lagged beta instruments is a significant predictor of the corresponding portfolio factor exposure. The regression R^2 for the low-R^2 (high-R^2) portfolio increases from 93.7% (98.7%) in the unconditional model to 96.5% (99.3%) in the conditional case. Modeling time variation in factor exposures thus explains a substantial portion of the remaining variation in portfolio returns. The annualized conditional alpha for the low-minus-high portfolio is 0.34%, which is insignificant at conventional levels. This lack of statistical significance in the conditional model is noteworthy because it comes in spite of the fact that test power is much higher in this specification. In particular, the portfolio alpha estimates are more precise in the conditional case because of the substantially improved model fit, such that the standard error of the conditional alpha is about two-thirds of the corresponding figure for the unconditional alpha. Finally, the reductions in alphas for the low-R^2 and low-minus-high R^2 portfolios are both large in economic magnitude and strongly significant, with p-values of 0.3% and 0.4%, respectively. Overall, the conditional models produce strong evidence that portfolio factor loadings are time varying, and modeling this time variation affects inferences about the R^2 strategies.

The remainder of Table III reports results for strategies formed on active share (Panel B), active weight (Panel C), volatility (Panel D), industry concentration (Panel E), and return gap (Panel F). The general conclusions from Panel A tend to carry over to these sets of test portfolios. The conditional specifications indicate that modeling time variation in portfolio factor loadings is important. Across the six sets of conditional models in Panels A-F, 40 of the 48 three-month lagged factor loading instruments are significant predictors at the 5% level. Allowing for time-varying factor loadings also produces an improvement in model fit for all of the strategy portfolios.

Moreover, introducing conditioning information is important for making inferences about fund performance for all six decile strategies in Panels A-F of Table III. The unconditional alpha of the active share strategy is significantly positive at the 10% level, the R^2, active weight, and industry concentration unconditional alphas are statistically significant at the 5% level, and the corresponding figures for the volatility and return gap portfolios are significant at the 1% level. In contrast, the conditional alphas for four of these six strategies—R^2, active share, volatility, and industry concentration—are statistically insignificant. The reductions in alpha for the six strategies are statistically significant at the 10% level with one exception—return gap (p-value of 12.9%). The return gap strategy in Panel F is the only one with a significant conditional performance measure at the 5% level, earning a conditional alpha of 1.42% (standard error of 0.39%) compared to the unconditional alpha estimate of 1.76% (standard error of 0.50%).

In addition to the evidence that conditional alphas are statistically weaker than their unconditional counterparts, the results in Table III have important economic implications. In particular, the conditional long-short alpha is less than the corresponding unconditional alpha in all cases, and the average percentage reduction in magnitude is 61%. These reductions in alpha are large in economic terms. For example, the 1.47% per year difference in unconditional performance for high- and low-AS mutual funds (Panel B) drops to just 0.07% after accounting for time-series variation in portfolio style, and the difference in unconditional and conditional alphas for the volatility strategy is a substantial 4.56% per year. We also find that evaluating performance with a conditional model is particularly important for the long leg of each strategy portfolio, as the conditional alpha is significantly less than the unconditional alpha at the 5% level for each of these portfolios. Moreover, each long leg portfolio has a negative conditional alpha estimate, indicating that the strategy-level abnormal returns are primarily attributable to significant negative performance in the short leg portfolios. Given that investors are unable to take short positions in mutual funds, this finding limits the practical appeal of activeness-related strategies.

Before proceeding, we note that the regression R^2s in Table III demonstrate that using lagged factor loading estimates as instruments for factor exposures substantially improves tracking performance.[18] A byproduct of this improvement is an increase in the precision of alpha estimates as unexplained return variance declines. In our tests, for example, the standard error of the active share strategy's conditional alpha in Panel B is 0.51% compared to 0.85% for the unconditional alpha. An increase in the precision of an alpha estimate leads to higher power to reject the null hypothesis of no performance.

Taken together, the results in Table III suggest that using conditional benchmarks is important for evaluating strategies that predict mutual fund performance and can have an economically meaningful impact on inferences. In particular, the conditional Carhart (1997) model results for the R^2, active share, volatility, and industry concentration strategies suggest that the primary driver of performance is not skill in security selection by mutual fund managers. Only the active weight and return gap measures remain significant predictors of abnormal fund performance relative to their conditional benchmarks. In Section 4.5, we revisit these results to further examine the potential sources of the unconditional alphas earned by the mutual fund strategies.

4.3 Lagged factor loading instrument design

Our conditional benchmarking approach relies on instruments based on lagged factor loadings to capture predictable, short-term movements in strategy factor exposures. As introduced in Section 3.2, we can estimate lagged factor loadings using either fund returns or the returns of fund holdings. Our base specifications use returns-based instruments when they are available for the entire sample period (i.e., for the portfolios sorted on lagged volatility) and holdings-based instruments otherwise. We also use three months of lagged data to calculate instruments in the base case. In this section, we examine the impact of these design choices on model fit and inferences about mutual fund performance. This analysis serves two purposes. First, we establish that our main empirical findings are robust to reasonable alternative specifications. Second, we provide recommendations for constructing instruments based on the fit of alternative conditional models.

[18] We test whether the improvements in tracking performance are statistically significant by comparing the unconditional and conditional regression adjusted R^2s in a bootstrap analysis. For each strategy, we form 25,000 bootstrap samples by drawing T monthly observations of strategy returns, factor returns, and strategy instruments with replacement, where T is the number of monthly observations in the strategy's sample period. We then estimate unconditional and conditional versions of the Carhart (1997) four-factor model using equations (4) and (5) for each bootstrap sample, calculate the adjusted R^2 for each model, and find the bootstrap p-value as the proportion of draws in which the unconditional R^2 is greater than the conditional R^2. The increase in R^2 from introducing conditioning information for each model in Table III is statistically significant at the 1% level.

We begin our analysis by studying tracking performance with alternative factor loading instruments. Figure 3 shows adjusted R^2s for the long leg portfolio of each strategy across several specifications. Given that the sample period with returns-based instruments is limited by data availability, Figure 3 plots R^2s for the January 1999-December 2015 period with the exception of a January 2000-December 2015 sample for the volatility strategy. Each panel shows R^2 values from an unconditional Carhart (1997) model and several conditional model specifications with returns-based or holdings-based instruments. The lagged factor loadings in the conditional models are estimated from daily data over the past one to 12 months or monthly data over the past 24 months.

Three basic patterns emerge from examining Figure 3. First, all of the conditional models exhibit improvements in adjusted R^2 relative to the corresponding unconditional model. Second, the conditional R^2s reflect a tradeoff for the instrument lag length. On the one hand, using more recent returns data allows the lagged loadings to better reflect the conditional exposures of stocks that are currently held in the mutual fund portfolios. On the other hand, estimating a Carhart (1997) regression model using a relatively longer time series of returns will reduce estimation error in the factor loadings. The three-month lagged factor loading instruments that we use in Section 4.2 provide a balance between these considerations and produce conditional benchmarks that generally perform well relative to the alternatives. Third, the R^2s for the returns-based and holdings-based instruments are similar for each specification, but there is some tendency for the models with returns-based lagged factor loadings to track strategy returns better compared to holdings-based specifications.

Table IV reports alpha estimates for alternative specifications of the conditional Carhart (1997) models for the six strategies of interest. The table presents long-short alpha estimates, the associated standard errors, and p-values from tests of whether the conditional alphas are less than their unconditional counterparts. For ease of comparison, Panel A shows unconditional alphas that correspond to the results in Panel B of Table I. Conditional model alpha estimates from specifications with lagged factor loading instruments from one-month, three-month, six-month, 12-month, and 24-month periods are displayed in Panel B. Panel C shows alphas from conditional models that combine three-month instruments with longer-term instruments from lagged 12-month or 24-month periods.

The results in Table IV show that our main findings from Section 4.2 are robust to alternative instruments. Beginning with the single-instrument specifications in Panel B, the five conditional models for the R^2 strategy produce alpha estimates between 0.34% and 0.79%. The conditional alphas are significantly lower than the unconditional alpha of 1.91% in all cases. The remaining strategies generally produce similar results to the base specification, with economically large reductions in estimated alphas after incorporating information from lagged factor loading instruments. Return gap is again an exception as it is a significant predictor of performance relative to any of the conditional models. In some cases, such as for the industry concentration strategy, the relatively poor fit of the conditional models with longer-term (i.e., 12-month or 24-month) lagged factor loading instruments observable in Figure 3 is accompanied by conditional alphas that are similar in magnitude to the unconditional alphas. This finding indicates that accurately modeling relatively short-term movements in strategy factor loadings is important for inferences. Finally, the two-instrument models in Panel C generally show little improvement in model fit (not shown) after adding either the 12-month or 24-month lagged factor loading instruments, and the conditional alpha estimates are similar in magnitude to those in the base specifications.

Overall, the results in this section show that our conclusions are not driven by a specific choice of lagged loading instruments. There are, however, systematic patterns in the tracking performance of conditional models with alternative instruments that should influence the design of conditional performance tests. Based on our analysis of the six strategies under consideration, our recommendations are to form returns-based instruments if this choice does not limit the sample period and to use relatively short (e.g., three-month) estimation periods to calculate lagged factor loading instruments.

4.4 Comparison to traditional instruments

We now compare the performance of our conditional benchmarking approach based on lagged factor loadings to the traditional methods in the literature that rely on macroeconomic predictors. We argue in Section 3 that the lagged loading instruments have desirable features for modeling portfolio exposures. In Table V, we investigate whether or not these instruments outperform the dividend yield, default spread, and term spread in tracking strategy returns. Panels A-F show results for the six strategies under consideration. For reference, each panel reproduces the results from Table III for the unconditional Carhart (1997) model and the conditional Carhart (1997) model with three-month lagged loadings as instruments.

In Panel A of Table V, Case C1 corresponds to a model that allows only the market factor loading to vary with conditioning information. That is, the market loading is specified as a linear function of the dividend yield, default spread, and term spread traditional state variables. This approach to instrumenting only for the market factor in a Carhart (1997) regression is common in the literature (e.g., Kacperczyk, Sialm, and Zheng (2005), Kosowski, Timmermann, Wermers, and White (2006), Huang, Sialm, and Zhang (2011), and Doshi, Elkamhi, and Simutin (2015)) and is motivated by the extensive evidence on the predictability of market returns using these variables. Using the three traditional instruments for market beta, the conditional alpha for the low-minus-high R^2 strategy only declines to 1.76% from the unconditional estimate of 1.91%. The regression R^2s for the low-R^2 portfolio and high-R^2 portfolio each exhibit a modest increase of between 0.0% and 0.3%. We next consider a conditional model (Case C2) that allows each of the four factor loadings to vary with the traditional instruments. This approach is adopted, for example, in Kacperczyk, Sialm, and Zheng (2008) and allows for considerably more flexibility than instrumenting for market beta alone. The conditional alpha for Case C2 of 0.94% is statistically significantly lower than the unconditional alpha with a p-value of 2.8%, but abnormal performance in Case C2 remains economically large in comparison to the 0.34% alpha from our conditional approach in Case C3. Finally, Case C4 combines the three-month lagged factor loading instruments with the traditional instruments, such that each of the four factors has four instruments. Results are similar across Cases C3 and C4, with conditional alpha estimates of 0.34% and 0.35%, respectively.

Panels B-F of Table V show the results for the remaining five strategies. Our specification in Case C3 produces better model fit compared to the methods using only traditional variables (Cases C1 and C2) for each portfolio. Additionally, Case C4 in each panel produces similar inferences to Case C3 for the abnormal performance of the six strategies, such that the traditional state variables have relatively little effect on the conditional models after including the lagged factor loading instruments. Thus, the results in Table V support the use of lagged portfolio factor loadings as superior instruments for factor exposures.

4.5 Decomposition and evaluation of strategy performance

Our main results in Section 4.2 show that, whereas the spreads in unconditional alphas are significantly positive for each of the six predictor variables, the corresponding conditional alphas are all substantially smaller in magnitude and only active weight (10% significance level) and return gap (1% significance level) remain as significant predictors of abnormal performance. Thus, the bulk of the unconditional performance of these mutual fund strategies is not likely to be attributable to skill in security selection by fund managers. In this section, we decompose the unconditional alphas of the strategy portfolios. As shown by Lewellen and Nagel (2006), Boguth, Carlson, Fisher, and Simutin (2011), and others, the difference between the unconditional and conditional alpha of a portfolio is a function of factor timing and volatility timing. In particular, systematic covariation between portfolio factor loadings and either the expected returns or volatilities of the factors can produce unconditional alphas that differ from conditional alphas.

For the Carhart (1997) model, the unconditional alpha estimate for a given portfolio can be decomposed as

$$\begin{aligned}
\hat{\alpha}_i^U = \overbrace{\hat{\alpha}_i^C}^{\text{Security selection}} &+ \overbrace{\text{cov}(\hat{\beta}_{i,t}^C, R_{MKT,t})}^{\text{Factor timing}} + \overbrace{(\bar{\beta}_{i,t}^C - \hat{\beta}_i^U)\bar{R}_{MKT,t}}^{\text{Volatility timing}} \\
&+ \text{cov}(\hat{s}_{i,t}^C, R_{SMB,t}) + (\bar{s}_{i,t}^C - \hat{s}_i^U)\bar{R}_{SMB,t} \\
&+ \text{cov}(\hat{h}_{i,t}^C, R_{HML,t}) + (\bar{h}_{i,t}^C - \hat{h}_i^U)\bar{R}_{HML,t} \\
&+ \text{cov}(\hat{u}_{i,t}^C, R_{UMD,t}) + (\bar{u}_{i,t}^C - \hat{u}_i^U)\bar{R}_{UMD,t}, \qquad (6)
\end{aligned}$$

where $\hat{\beta}_{i,t}^C$ is the fitted conditional loading on the market factor, $\bar{\beta}_{i,t}^C$ is the average conditional loading, $\hat{\beta}_i^U$ is the unconditional market loading estimate, and the terms for the remaining three factors are defined analogously. For each of the four factors, a direct factor timing term and a factor bias effect term can contribute to differences between the unconditional and conditional portfolio alpha. The direct factor timing terms measure the covariances between factor loadings and factor returns. A positive covariance between a portfolio's exposure to a factor and the factor's realized return will have a positive effect on the measured unconditional alpha. The factor bias terms reflect the differences between the average conditional factor exposures and the unconditional loadings. These terms are related to volatility timing. For example, if the conditional loading on the market factor for a portfolio tends to be high when the market factor is highly volatile, then the portfolio's unconditional market factor loading will overstate its average conditional exposure to the market factor.

Our proposed method for estimating conditional alpha combined with this decomposition of unconditional alpha provides an approach for attribution analysis of a predictor of mutual fund performance. A given strategy's conditional alpha may capture the security selection skill of the managers of mutual funds held in the strategy portfolio.[19] The remainder of the unconditional performance of a strategy is attributable to factor timing or volatility timing, which could potentially, but not necessarily, be indicative of managerial skill as discussed further below.

[19]As noted by Ferson and Mo (2016) and others, the conditional alpha could also reflect managerial ability to execute low cost trades or manage an efficient securities lending operation.

Table VI shows results from empirical decompositions of the unconditional alphas for the strategy long-short portfolios. The conditional models in this table correspond to the base results from Table III. For each strategy, Panel B reports the total factor timing and total volatility timing effects as well as the factor timing components associated with each factor. We evaluate the statistical significance of each component using a bootstrap approach. Specifically, we draw monthly observations with replacement from the sample observations to create 25,000 bootstrapped samples for each strategy that match the original sample length. For each monthly draw, a full set of portfolio returns, instruments, and factor returns is taken from a single month to preserve the structure of the data. Given a bootstrapped sample, we estimate unconditional and conditional Carhart (1997) model regressions and calculate the market timing and volatility timing components as described above. A * (**) [***] in Table VI indicates that the 90% (95%) [99%] bootstrap confidence interval of the corresponding estimate does not include zero.

The results in Table VI indicate that factor timing is the key driver of the differences between unconditional and conditional alpha estimates. In particular, the total factor timing effects are positive for all six strategies. Moreover, the unconditional alpha estimates are influenced by statistically significant factor timing for the strategies based on R^2 (factor timing effects contribute 1.82% of the 1.91% unconditional alpha estimate), active weight (0.75% of 1.40%), volatility (4.47% of 6.17%), and return gap (0.48% of 1.76%). Significantly positive factor timing effects occur for the market factor (volatility strategy) and value factor (R^2, active weight, and return gap strategies). Factor timing effects also make substantial positive contributions to the unconditional alpha estimates of the active share (1.08% of 1.47%) and industry concentration (0.50% of 1.37%) strategies, but these effects are statistically insignificant. Volatility timing, on the other hand, tends to have a smaller economic impact on strategy alphas. Overall, the results in Table VI suggest that factor timing, rather than security selection, is the primary driver of the positive unconditional performance estimates for most of the managerial activeness measures that we consider.

Positive factor timing effects at the strategy level may or may not be attributable to mutual fund manager skill, as timing effects have three potential sources. First, the strategy may tend to invest in managers with the skill to time factors based on information or trading rules that are not publicly known. Second, the strategy may identify mutual fund managers who mechanically follow known timing strategies based on publicly available information. Third, the strategy may shift its investments across mutual funds with different factor exposures when the portfolios are rebalanced such that the strategy's factor exposures change. In our view, only the first of these scenarios lends itself to an interpretation of the portfolio sorting characteristic as an indicator of managerial skill.

We proceed to investigate whether the observed factor timing effects arise at the strategy level or fund level. As previously discussed, the strategy portfolios are rebalanced monthly and exhibit substantial turnover. The changing composition of the portfolios produces challenges when determining whether factor timing effects are generated by the funds held within the strategy portfolios or by strategy turnover across funds, because a relatively long sample period is necessary to establish the presence of timing ability. To overcome this issue, our approach is to measure the total factor timing effects over the full sample period for every cohort portfolio generated by a given strategy. In particular, each strategy has T rebalance dates on which portfolios are formed, where T is the number of months in the sample period. We form each of these T long-short portfolios, calculate the portfolio returns for the entire sample period, and estimate the total factor timing effect from equation (6). This procedure generates T cohort timing effects associated with each strategy. If a given proxy for managerial activeness successfully identifies funds with positive timing ability and timing skill is persistent, then the cohort portfolios should systematically exhibit positive factor timing effects.

Figure 4 shows histograms of the factor timing effects for cohort portfolios in percent per year. Each panel also plots the total factor timing effect for the strategy portfolio (dashed line) from Table VI. Across the six strategies, the cohort portfolios generally display little factor timing compared to the strategy portfolios. The only managerial activeness measures for which any of the T cohort portfolios generates a larger factor timing effect than the strategy portfolio are industry concentration (two of 312 cohort portfolios) and return gap (five of 312 cohort portfolios). For the return gap measure, the distribution of cohort timing is nonetheless centered around zero.

Panel C of Table VI provides summary statistics for the total factor timing effects across the cohort portfolios. The average cohort timing effects are small compared to the corresponding strategy figure for each of the strategies with significant timing effects. In particular, the average cohort timing effects are small compared to strategy timing for the R^2 (0.46% average for the cohorts versus 1.82% for the strategy), active weight (0.13% versus 0.75%), volatility (0.73% versus 4.47%), and return gap (−0.02% versus 0.48%) portfolios. Overall, the results in Figure 4 and Table VI indicate that, whereas the strategy portfolios tend to generate substantial positive factor timing effects, the cohort portfolios generated by the strategies almost never produce these large effects. These findings are consistent with the view that strategy-level turnover across funds accounts for most of the factor timing effects over the sample period.

Our final analysis provides additional evidence about timing at the strategy versus fund level by concentrating on strategy holdings and performance around the "technology bubble" period. We demonstrate in Section 4.1 that the performance of several of the strategies is concentrated around this time. Periods characterized by volatile factor returns, exemplified by the 1998-2000 episode, provide the greatest potential to produce large factor timing effects.

To examine the mutual fund strategies during this period, we follow Brunnermeier and Nagel (2004) and measure the exposure of the strategy portfolios to technology stocks over the period January 1998-December 2000. We begin by ranking Nasdaq stocks by their price-to-sales ratio, where sales data are from the Compustat Fundamentals Annual File and lagged at least six months. We then calculate the weight in each mutual fund that is held in Nasdaq stocks that rank in the top quintile of price-to-sales using fund holdings data. Finally, we find the average fund weight in technology stocks for both the long leg and short leg portfolios of each strategy, and the difference between these two weights proxies for the net exposure of the strategy to technology stocks.

Figure 5 plots the monthly exposure to technology stocks for each strategy over the January 1998-December 2000 period. A positive (negative) number indicates that the long (short) leg strategy portfolio has a larger technology exposure. We also show the net exposure for buy-and-hold versions of the strategy portfolios formed at the beginning of January 1998. The dashed line indicates the timing of the Nasdaq peak in March 2000.

The times series plots of technology holdings shown in Figure 5 indicate that the strategies collectively navigated through the technology bubble quite well. The R^2 strategy, for example, maintained a positive net exposure in technology stocks until March 2000 (the month of the peak) and subsequently had a negative exposure during the crash months. This shift in exposure could arise either from strategy-level turnover or from changes in fund exposures. The January 1998 cohort portfolio for the R^2 strategy exhibits relatively little time variation in technology exposure, indicating that turnover of funds at the strategy level is largely responsible for the changes in technology weights.

Results for the active share, active weight, industry concentration, and return gap strategies in Figure 5 each show noticeable declines in technology exposures within three months of the peak. In contrast, the January 1998 cohort portfolio for each strategy does not follow a similar pattern. The volatility strategy maintains a very sizeable negative exposure to technology stocks over the January 2000-December 2000 period, but the strategy portfolio avoids any effects of being short technology stocks in the pre-2000 period due to data availability. The results in Figure 5 are consistent with positive factor timing effects that arise from strategy-level turnover rather than fund-level timing ability.

In sum, decompositions of the unconditional performance of the six managerial activeness strategies produce relatively limited support of predictable abnormal performance in mutual fund returns. Four of the strategies have insignificant conditional alphas, such that there is little evidence of managerial skill in security selection associated with these measures. Only active weight and return gap predict differential security selection ability, but these effects are driven by the statistically significant negative conditional net-of-fee alphas of the short legs of the strategy portfolios. Significant factor timing effects strongly contribute to the positive unconditional alphas of the R^2, active weight, volatility, and return gap strategies. The positive factor timing effects, however, appear to be attributable to strategy-level rebalancing across funds rather than managerial timing ability.

5 Conclusion

In this paper, we show that the conventional approach to evaluating portfolios of mutual funds based on unconditional factor-model regressions is problematic. Specifically, these portfolios often exhibit pronounced jumps in style exposures and predictable trends in factor loadings over time. Evidence of managerial skill in such instances has the potential to be contaminated by a poorly specified benchmark model that fails to account for changes in the portfolios' style exposures. We introduce a method for evaluating conditional portfolio performance that builds on recent innovations from the asset pricing literature. This approach successfully incorporates information from lagged mutual fund factor exposures in assessing managerial skill in security selection, factor timing, and volatility timing.

Among a broad set of six strategies based on managerial activeness, we find that model fit substantially improves after allowing for time variation in factor loadings. Further, the economic and statistical evidence of abnormal performance is considerably reduced after incorporating conditioning information. In particular, whereas all six of the strategies earn significant unconditional alphas, the conditional strategy alphas are reduced in magnitude by an average of 61% relative to the unconditional alphas, and only two measures (active weight and return gap) produce statistically significant alphas relative to conditional benchmarks. The unconditional abnormal performance for several of the strategies is attributable to positive outcomes of the strategies' implicit style bets over relatively short periods within the sample. Further, much of the realized strategy-level performance related to security selection and factor timing is driven by poor performance of low-activeness funds (i.e., the short leg of each strategy). The evidence collectively suggests that proxies for fund activeness are not reliable indicators of managerial skill.

Our proposed modeling approach should be a useful tool for evaluating similar evidence in past and future research on managed portfolios. For example, the unconditional performance of strategies based on manager characteristics and trading styles (e.g., Cohen, Coval, and Pástor (2005), Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014), and Wei, Wermers, and Yao (2015)) or contractual incentive structures (e.g., Massa and Patgiri (2009) and Huang, Sialm, and Zhang (2011)) could be decomposed using our conditional framework. Further, our method outperforms alternative approaches used in the literature, and the improvements in return tracking translate to increased statistical power to identify skilled managers.

References

Amihud, Yakov, and Ruslan Goyenko, 2013, Mutual fund's R^2 as predictor of performance, *Review of Financial Studies* 26, 667–694.

Avramov, Doron, and Tarun Chordia, 2006, Predicting stock returns, *Journal of Financial Economics* 82, 387–415.

Becker, Connie, Wayne Ferson, David H. Myers, and Michael J. Schill, 1999, Conditional market timing with benchmark investors, *Journal of Financial Economics* 52, 119–148.

Berk, Jonathan B., and Jules H. van Binsbergen, 2015, Measuring skill in the mututal fund industry, *Journal of Financial Economics* 118, 1–20.

Boguth, Oliver, Murray Carlson, Adlai Fisher, and Mikhail Simutin, 2011, Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas, *Journal of Financial Economics* 102, 363–389.

Bollen, Nicolas P.B., and Robert E. Whaley, 2009, Hedge fund risk dynamics: Implications for performance appraisal, *Journal of Finance* 64, 985–1035.

Brunnermeier, Markus K., and Stefan Nagel, 2004, Hedge funds and the technology bubble, *Journal of Finance* 59, 2013–2040.

Busse, Jeffrey A., 1999, Volatility timing in mutual funds: Evidence from daily returns, *Review of Financial Studies* 12, 1009–1041.

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

Chan, K. C., 1988, On the contrarian investment strategy, *Journal of Business* 61, 147–163.

Cohen, Randolph B., Joshua D. Coval, and L'uboš Pástor, 2005, Judging fund managers by the company they keep, *Journal of Finance* 60, 1057–1096.

Cooper, Michael J., and Stefano Gubellini, 2011, The critical role of conditioning information in determining if value is riskier than growth, *Journal of Empirical Finance* 18, 289–305.

Cremers, K. J. Martijn, and Ankur Pareek, 2016, Patient capital outperformance: The investment skill of high active share managers who trade infrequently, *Journal of Financial Economics* 122, 288–306.

Cremers, K. J. Martijn, and Antti Petajisto, 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic based benchmarks, *Journal of Finance* 52, 1035–1058.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197–226.

Doshi, Hitesh, Redouane Elkamhi, and Mikhail Simutin, 2015, Managerial activeness and mutual fund performance, *Review of Asset Pricing Studies* 5, 156–184.

Evans, Richard B., 2010, Mutual fund incubation, *Journal of Finance* 65, 1581–1611.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

——— , 1993, Common risk factors in the returns on bonds and stocks, *Journal of Financial Economics* 33, 3–53.

Ferson, Wayne E., 2010, Investment Performance Evaluation, *Annual Reviews of Financial Economics* 2, 207–234.

——— , 2013, Conditional performance evaluation, in Cheng-Few Lee, and Alice C. Lee, eds.: *Encyclopedia of Finance* (Springer), United States, chap. 11, pp. 279–285.

Ferson, Wayne E., and Haitao Mo, 2016, Performance measurement with selectivity, market and volatility timing, *Journal of Financial Economics* 121, 93–110.

Ferson, Wayne E., Sergei Sarkissian, and Timothy Simin, 2008, Asset pricing models with conditional betas and alphas: The effects of data snooping and spurious regression, *Journal of Financial and Quantitative Analysis* 43, 331–354.

Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.

Grant, Dwight, 1977, Portfolio performance and the "cost" of timing decisions, *Journal of Finance* 32, 837–846.

Grundy, Bruce D., and J. Spencer Martin, 2001, Understanding the nature and the risks of the source of the rewards to momentum investing, *Review of Financial Studies* 14, 29–78.

Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Henriksson, Roy D., 1984, Market timing and mutual fund performance: An empirical investigation, *Journal of Business* 57, 73–96.

Henriksson, Roy D., and Robert C. Merton, 1981, On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513–533.

Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang, 2011, Risk shifting and mutual fund performance, *Review of Financial Studies* 24, 2575–2616.

Jagannathan, Ravi, and Robert A. Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 59, 217–235.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jiang, George J., Tong Yao, and Tong Yu, 2007, Do mutual funds time the market? Evidence from portfolio holdings, *Journal of Financial Economics* 86, 724–758.

Jordan, Bradford D., and Timothy B. Riley, 2015, Volatility and mutual fund manager skill, *Journal of Financial Economics* 118, 289–298.

Kacperczyk, Marcin, and Amit Seru, 2007, Fund manager use of public information: New evidence on managerial skills, *Journal of Finance* 62, 485–528.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance* 60, 1983–2011.

——— , 2008, Unobserved actions of mutual funds, *Review of Financial Studies* 21, 2379–2416.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2014, Time-varying fund manager skill, *Journal of Finance* 69, 1455–1484.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* 61, 2551–2595.

Lewellen, Jonathan, and Stefan Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289–314.

Mamaysky, Harry, Matthew I. Spiegel, and Hong Zhang, 2008, Estimating the dynamics of mutual fund alphas and betas, *Review of Financial Studies* 21, 233–264.

Massa, Massimo, and Rajdeep Patgiri, 2009, Incentives and mutual fund performance: Higher performance or just higher risk taking?, *Review of Financial Studies* 22, 1777–1815.

Patton, Andrew, and Tarun Ramadorai, 2013, On the high-frequency dynamics of hedge fund risk exposures, *Journal of Finance* 68, 597–635.

Petajisto, Antti, 2013, Active share and mutual fund performance, *Financial Analysts Journal* 69, 79–93.

Sharpe, William F., 1992, Asset allocation: Management style and performance measurement, *Journal of Portfolio Management* 18, 7–19.

Sun, Zheng, Ashley Wang, and Lu Zheng, 2012, The road less traveled: Strategy distinctiveness and hedge fund performance, *Review of Financial Studies* 25, 96–143.

Titman, Sheridan, and Cristian Tiu, 2011, Do the best hedge funds hedge?, *Review of Financial Studies* 24, 123–168.

Treynor, Jack L., and Kay K. Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131–136.

Wei, Kelsey D., Russ Wermers, and Tong Yao, 2015, Uncommon value: The characteristics and investment performance of contrarian funds, *Management Science* 61, 2394–2414.

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

Table I: Unconditional four-factor regressions.
The table reports unconditional Carhart (1997) four-factor regression results for decile portfolios sorted on lagged R^2, active share (AS), active weight (AW), volatility (Vol), industry concentration (ICI), and return gap ($RetGap$). Panel A presents results corresponding to the sample period from the original study on each predictor variable, and Panel B extends the sample periods in Panel A through December 2015. The portfolios are equal weighted and rebalanced monthly. The unconditional alpha estimates (α_i^U) are reported in percentage per year, and the numbers in parentheses are White (1980) standard errors. For the alpha estimates, ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively, using a two-tailed test.

	Low−High R^2	High−Low AS	High−Low AW	Low−High Vol	High−Low ICI	High−Low $RetGap$
Panel A: Original sample period						
Sample start date	1990:01	1990:01	1980:04	2000:01	1984:01	1984:01
Sample end date	2010:12	2003:12	2013:12	2013:12	1999:12	2003:12
Alpha, α_i^U (%)	2.56***	1.10	1.38**	6.21***	1.49*	1.81**
	(0.96)	(1.28)	(0.62)	(2.08)	(0.82)	(0.73)
Panel B: Extended sample period						
Sample start date	1990:01	1990:01	1980:04	2000:01	1984:01	1984:01
Sample end date	2015:12	2015:12	2015:12	2015:12	2015:12	2015:12
Alpha, α_i^U (%)	1.91**	1.47*	1.40**	6.17***	1.37**	1.76***
	(0.84)	(0.85)	(0.60)	(1.93)	(0.63)	(0.50)
Factor loadings						
R_{MKT} loading	−0.07	0.00	−0.01	−0.53	0.08	0.04
R_{SMB} loading	0.26	0.61	0.05	−0.65	0.36	0.14
R_{HML} loading	0.13	0.24	−0.01	0.54	−0.21	−0.03
R_{UMD} loading	0.05	−0.05	−0.01	0.01	−0.01	0.11

Table II: Portfolio characteristics.
The table reports summary statistics for decile portfolios sorted on lagged R^2, active share (AS), active weight (AW), volatility (Vol), industry concentration (ICI), and return gap ($RetGap$). For each set of portfolios, "Low" ("High") represents the funds with the lowest (highest) values of the sorting variable. Panel A presents average net excess return, average gross excess return, and standard deviation of net return in percentage per year. Panel B reports properties of the mutual funds contained in each portfolio. "TNA" is total net assets, "Expense ratio" is the ratio of total investment that shareholders pay for the fund's operating expenses to assets under management, and "Turnover" is the minimum of aggregated sales or purchases of securities divided by the average 12-month TNA of the fund. The figures in Panel B are time-series averages of the monthly cross-sectional average characteristics for each portfolio. Panel C reports annualized portfolio turnover. Monthly turnover is computed as 0.5 times the sum of the absolute values of the change in portfolio weights in each underlying mutual fund. The annual turnover figures are computed by multiplying the monthly turnover values by 12. Panel D presents the time-series standard deviation of the equal-weighted lagged factor loadings, estimated from Carhart (1997) model regressions using the prior three months of daily data. The lagged loadings are estimated from mutual fund returns for the volatility portfolios and from mutual fund holdings for the other portfolios.

	R^2		AS		AW		Vol		ICI		$RetGap$	
	Low	High	High	Low	High	Low	Low	High	High	Low	High	Low
Panel A: Properties of portfolio excess returns (annualized)												
Mean (Net, %)	8.35	5.81	8.46	5.77	8.22	6.91	5.96	2.01	7.78	6.33	8.51	5.71
Mean (Gross, %)	9.73	6.84	9.90	6.79	9.50	8.02	7.16	3.34	9.12	7.33	9.85	6.98
Std. dev. (Net, %)	14.97	15.13	17.13	14.56	16.74	16.62	12.31	24.91	18.07	14.79	18.21	17.03
Panel B: Properties of underlying mutual funds												
TNA ($BB)	0.92	2.07	0.45	2.84	0.70	0.75	2.85	0.87	0.84	1.47	0.82	0.93
Expense ratio (%)	1.39	1.02	1.45	1.01	1.29	1.12	1.20	1.34	1.34	0.99	1.35	1.27
Turnover (%)	106	68	79	67	91	74	61	114	95	73	118	110
Panel C: Portfolio turnover (annualized)												
Turnover (%)	135	147	57	52	116	109	95	93	76	112	258	294
Panel D: Standard deviation of formation-period factor loadings												
R_{MKT} loading	0.09	0.03	0.11	0.03	0.09	0.07	0.05	0.11	0.12	0.03	0.11	0.09
R_{SMB} loading	0.13	0.12	0.11	0.05	0.16	0.14	0.16	0.15	0.15	0.06	0.18	0.11
R_{HML} loading	0.18	0.08	0.23	0.07	0.17	0.15	0.19	0.23	0.20	0.08	0.18	0.16
R_{UMD} loading	0.13	0.06	0.12	0.03	0.13	0.10	0.09	0.30	0.20	0.04	0.15	0.13

Table III: Conditional benchmark models.

The table reports Carhart (1997) four-factor model regression results for decile portfolios sorted on lagged R^2 (Panel A), active share (Panel B), active weight (Panel C), volatility (Panel D), industry concentration (Panel E), and return gap (Panel F). The return regression is given by $R_{i,t} = \alpha_i^C + (\lambda_{i,0} + \lambda_{i,1} Z_{i,t-1}^{MKT}) R_{MKT,t} + (\gamma_{i,0} + \gamma_{i,1} Z_{i,t-1}^{SMB}) R_{SMB,t} + (\eta_{i,0} + \eta_{i,1} Z_{i,t-1}^{HML}) R_{HML,t} + (\nu_{i,0} + \nu_{i,1} Z_{i,t-1}^{UMD}) R_{UMD,t} + \varepsilon_{i,t}$. The unconditional models correspond to regressions with no instruments for the factor loadings. In each conditional model, the conditioning variables ($Z_{i,t-1}^k$) for a given portfolio are the three-month lagged loadings for the corresponding factors. We use instruments based on mutual fund returns for the volatility portfolios in Panel D and instruments based on mutual fund holdings in the other panels. The estimates of α_i^U and α_i^C are reported in percentage per year, and the numbers in parentheses are White (1980) standard errors. For each regression, R^2 is the adjusted R^2 value. Each reported p-value ($p(\alpha_i^C = \alpha_i^U)$) is for a one-sided test of the null hypothesis that the conditional portfolio alpha is equal to the corresponding unconditional alpha against the alternative that the conditional alpha is less than the unconditional alpha. For the alpha estimates, ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively, using a two-tailed test. The sample period for each panel is the corresponding extended sample period in Panel B of Table I.

	Unconditional model		Conditional model										
						$R_{MKT,t}\times$		$R_{SMB,t}\times$		$R_{HML,t}\times$		$R_{UMD,t}\times$	
Decile	α_i^U	R^2	α_i^C	R^2	$p(\alpha_i^C = \alpha_i^U)$	1	β^{L3}	1	s^{L3}	1	h^{L3}	1	u^{L3}
Panel A: R^2 portfolios													
Low	0.49	93.7	-1.03^{*}	96.5	0.003	0.53	0.41	0.08	0.53	0.02	0.75	0.03	0.68
	(0.81)		(0.60)			(0.22)	(0.22)	(0.04)	(0.12)	(0.02)	(0.08)	(0.01)	(0.08)
High	-1.42^{***}	98.7	-1.37^{***}	99.3	—	0.44	0.54	0.00	0.90	0.00	0.60	0.00	0.52
	(0.35)		(0.25)			(0.19)	(0.19)	(0.01)	(0.08)	(0.01)	(0.12)	(0.01)	(0.17)
Low−High	1.91^{**}	—	0.34	—	0.004								
	(0.84)		(0.55)										
Panel B: Active share portfolios													
High	0.29	94.9	-1.25^{**}	97.7	0.007	0.45	0.53	0.32	0.39	−0.02	0.94	−0.01	0.65
	(0.80)		(0.51)			(0.17)	(0.17)	(0.10)	(0.15)	(0.02)	(0.07)	(0.01)	(0.09)
Low	-1.18^{***}	99.3	-1.32^{***}	99.4	—	0.34	0.63	−0.04	0.44	−0.01	0.31	0.00	0.48
	(0.25)		(0.23)			(0.20)	(0.19)	(0.02)	(0.21)	(0.01)	(0.15)	(0.00)	(0.13)
High−Low	1.47^{*}	—	0.07	—	0.012								
	(0.85)		(0.51)										

(*continued*)

Table III—*Continued*

	Unconditional model		Conditional model										
						$R_{MKT,t}\times$		$R_{SMB,t}\times$		$R_{HML,t}\times$		$R_{UMD,t}\times$	
Decile	α_i^U	R^2	α_i^C	R^2	$p(\alpha_i^C = \alpha_i^U)$	1	β^{L3}	1	s^{L3}	1	h^{L3}	1	u^{L3}
Panel C: Active weight portfolios													
High	−0.17	96.7	−0.78*	98.0	0.047	0.67	0.29	0.09	0.55	−0.01	0.57	0.01	0.69
	(0.52)		(0.41)			(0.11)	(0.10)	(0.04)	(0.08)	(0.02)	(0.07)	(0.01)	(0.06)
Low	−1.57***	96.4	−1.45***	98.3	—	0.83	0.14	−0.05	1.00	−0.02	0.70	0.02	0.57
	(0.59)		(0.41)			(0.12)	(0.11)	(0.04)	(0.09)	(0.01)	(0.11)	(0.01)	(0.14)
High−Low	1.40**	—	0.67*	—	0.054								
	(0.60)		(0.38)										
Panel D: Volatility portfolios													
Low	2.03**	93.5	−0.08	97.8	0.001	0.64	0.27	0.00	0.75	0.03	0.98	0.00	0.74
	(0.80)		(0.47)			(0.15)	(0.18)	(0.02)	(0.09)	(0.01)	(0.07)	(0.01)	(0.17)
High	−4.15***	94.0	−1.69*	97.7	—	−0.30	1.30	0.03	0.82	−0.03	0.80	0.03	0.83
	(1.46)		(0.95)			(0.38)	(0.34)	(0.11)	(0.19)	(0.03)	(0.20)	(0.02)	(0.07)
Low−High	6.17***	—	1.61	—	0.003								
	(1.93)		(1.11)										
Panel E: Industry concentration portfolios													
High	0.13	96.2	−0.50	97.2	0.045	0.76	0.23	0.24	0.16	−0.07	0.53	0.01	0.56
	(0.65)		(0.54)			(0.13)	(0.12)	(0.06)	(0.13)	(0.03)	(0.09)	(0.01)	(0.06)
Low	−1.24***	99.3	−1.26***	99.4	—	0.79	0.18	−0.01	0.58	0.04	0.21	0.00	0.14
	(0.23)		(0.22)			(0.19)	(0.18)	(0.01)	(0.16)	(0.01)	(0.11)	(0.00)	(0.12)
High−Low	1.37**	—	0.76	—	0.042								
	(0.63)		(0.52)										
Panel F: Return gap portfolios													
High	−0.10	96.4	−0.71	97.4	0.033	0.46	0.50	0.21	0.41	−0.05	0.34	0.06	0.66
	(0.62)		(0.55)			(0.15)	(0.13)	(0.06)	(0.12)	(0.02)	(0.08)	(0.02)	(0.10)
Low	−1.86***	97.1	−2.13***	97.8	—	0.50	0.45	0.12	0.42	−0.03	0.60	−0.01	0.61
	(0.56)		(0.47)			(0.17)	(0.16)	(0.05)	(0.15)	(0.02)	(0.11)	(0.01)	(0.10)
High−Low	1.76***	—	1.42***	—	0.129								
	(0.50)		(0.39)										

Table IV: Comparison of lagged factor loading instruments.
The table reports conditional Carhart (1997) four-factor model regression results for decile portfolios sorted on lagged R^2, active share (AS), active weight (AW), volatility (Vol), industry concentration (ICI), and return gap ($RetGap$). The return regression is given by $R_{i,t} = \alpha_i^C + (\lambda_{i,0} + \lambda'_{i,1} Z_{i,t-1}^{MKT}) R_{MKT,t} + (\gamma_{i,0} + \gamma'_{i,1} Z_{i,t-1}^{SMB}) R_{SMB,t} + (\eta_{i,0} + \eta'_{i,1} Z_{i,t-1}^{HML}) R_{HML,t} + (\nu_{i,0} + \nu'_{i,1} Z_{i,t-1}^{UMD}) R_{UMD,t} + \varepsilon_{i,t}$. The results in Panel A correspond to unconditional models with no instruments for the factor loadings. Panel B presents results for conditional models, in which the conditioning variables for a given portfolio are the one-, three-, six-, 12-, or 24-month lagged loadings for the corresponding factors. Panel C presents results for conditional models, in which the conditioning variables for a given portfolio are the three-month lagged loadings and either the 12- or 24-month lagged loadings for the corresponding factors. The 24-month instruments are estimated from monthly four-factor regressions, and all other instruments are based on daily four-factor regressions. For each model, the table shows the estimate of the long-short portfolio alpha (α_Δ^C) in percentage per year and the corresponding White (1980) standard error ($\sigma(\alpha_\Delta^C)$). For each conditional model in Panels B and C, the table also reports a p-value (p-val.) for the one-sided test of the null hypothesis that the conditional alpha is equal to the corresponding unconditional alpha from Panel A against the alternative that the conditional alpha is less than the corresponding unconditional alpha. For the alpha estimates, ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively, using a two-tailed test.

	Low−High R^2			High−Low AS			High−Low AW			Low−High Vol			High−Low ICI			High−Low $RetGap$		
Instruments	α_Δ^C	$\sigma(\alpha_\Delta^C)$	p-val.	α_Δ^C	$\sigma(\alpha_\Delta^C)$	p-val.	α_Δ^C	$\sigma(\alpha_\Delta^C)$	p-val.	α_Δ^C	$\sigma(\alpha_\Delta^C)$	p-val.	α_Δ^C	$\sigma(\alpha_\Delta^C)$	p-val.	α_Δ^C	$\sigma(\alpha_\Delta^C)$	p-val.
Panel A: Unconditional model																		
None	1.91**	0.84	—	1.47*	0.85	—	1.40**	0.60	—	6.17***	1.93	—	1.37**	0.63	—	1.76***	0.50	—
Panel B: Conditional models with one instrument for each loading																		
L1	0.74	0.55	0.02	0.18	0.60	0.01	1.01***	0.38	0.17	2.27*	1.25	0.01	0.86	0.54	0.06	1.32***	0.43	0.07
L3	0.34	0.55	0.00	0.07	0.51	0.01	0.67*	0.38	0.05	1.61	1.11	0.00	0.76	0.52	0.04	1.42***	0.39	0.13
L6	0.53	0.55	0.01	0.33	0.53	0.03	0.57	0.38	0.03	1.66	1.24	0.00	1.05*	0.54	0.17	1.33***	0.40	0.07
L12	0.78	0.58	0.02	0.45	0.55	0.04	0.55	0.38	0.03	1.76	1.39	0.00	1.16**	0.59	0.25	1.41***	0.40	0.09
L24	0.79	0.64	0.01	0.08	0.58	0.01	0.70*	0.42	0.04	2.61*	1.38	0.00	1.34**	0.60	0.47	1.66***	0.44	0.35
Panel C: Conditional models with two instruments for each loading																		
L3 & L12	0.43	0.54	0.01	0.05	0.51	0.01	0.73**	0.37	0.07	1.34	1.11	0.00	1.06**	0.53	0.20	1.58***	0.39	0.29
L3 & L24	0.23	0.55	0.00	−0.25	0.50	0.01	0.60*	0.36	0.05	1.71*	1.03	0.00	0.81	0.52	0.06	1.63***	0.38	0.35

Table V: Comparison of lagged factor loadings and traditional instruments in conditional benchmark models.

The table reports conditional Carhart (1997) four-factor model regression results for decile portfolios sorted on lagged R^2 (Panel A), active share (Panel B), active weight (Panel C), volatility (Panel D), industry concentration (Panel E), and return gap (Panel F). The return regression is given by $R_{i,t} = \alpha_i^C + (\lambda_{i,0} + \lambda'_{i,1} Z_{i,t-1}^{MKT}) R_{MKT,t} + (\gamma_{i,0} + \gamma'_{i,1} Z_{i,t-1}^{SMB}) R_{SMB,t} + (\eta_{i,0} + \eta'_{i,1} Z_{i,t-1}^{HML}) R_{HML,t} + (\nu_{i,0} + \nu'_{i,1} Z_{i,t-1}^{UMD}) R_{UMD,t} + \varepsilon_{i,t}$. The conditioning variables, $Z_{i,t-1}^k$, for a given portfolio include traditional instruments (i.e., dividend yield, default spread, and term spread) and three-month lagged factor loadings ("L3"). In each panel, we present results for unconditional models with no instruments for the factor loadings (Case "U"), conditional models with traditional instruments for the market factor (Case "C1"), conditional models with traditional instruments for each factor (Case "C2"), conditional models with lagged beta instruments for each factor (Case "C3"), and conditional models with lagged beta instruments and traditional instruments for each factor (Case "C4"). For each set of portfolios, "L" represents the long leg based on the given sorting variable, "S" represents the short leg, and "Δ" represents the hypothetical difference portfolio. The estimates of α_L^C, α_S^C, and α_Δ^C are reported in percentage per year, and $\sigma(\alpha_\Delta^C)$ is the White (1980) standard error for the corresponding difference in conditional alphas. R_L^2 and R_S^2 are the adjusted R^2 values for the long and short portfolios, respectively. For each conditional model, the table reports a p-value ($p(\alpha_\Delta^C = \alpha_\Delta^U)$) for the one-sided test of the null hypothesis that the difference in conditional alphas is equal to the corresponding difference in unconditional alphas against the alternative that the difference in conditional alphas is less than the difference in unconditional alphas. For the alpha estimates, ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively, using a two-tailed test.

			Parameter estimates					Model fit	
Case	Instruments	Loading(s)	α_L^C	α_S^C	α_Δ^C	$\sigma(\alpha_\Delta^C)$	$p(\alpha_\Delta^C = \alpha_\Delta^U)$	R_L^2	R_S^2
			Panel A: R^2 portfolios						
U	None		0.49	−1.42***	1.91**	0.84	—	93.7	98.7
C1	Traditional	Market	0.35	−1.41***	1.76**	0.79	0.248	94.0	98.7
C2	Traditional	All	−0.57	−1.51***	0.94	0.75	0.028	95.4	98.8
C3	L3	All	−1.03*	−1.37***	0.34	0.55	0.004	96.5	99.3
C4	L3 & Traditional	All	−1.02*	−1.37***	0.35	0.52	0.007	96.8	99.4
			Panel B: Active share portfolios						
U	None		0.29	−1.18***	1.47*	0.85	—	94.9	99.3
C1	Traditional	Market	−0.23	−1.06***	0.83	0.74	0.072	95.5	99.4
C2	Traditional	All	−1.11**	−1.13***	0.01	0.59	0.014	97.4	99.4
C3	L3	All	−1.25**	−1.32***	0.07	0.51	0.012	97.7	99.4
C4	L3 & Traditional	All	−1.38***	−1.19***	−0.19	0.52	0.009	97.9	99.5
			Panel C: Active weight portfolios						
U	None		−0.17	−1.57***	1.40**	0.60	—	96.7	96.4
C1	Traditional	Market	−0.26	−1.65***	1.39**	0.63	0.443	96.7	96.4
C2	Traditional	All	−0.34	−1.68***	1.34**	0.67	0.414	97.3	97.6
C3	L3	All	−0.78*	−1.45***	0.67*	0.38	0.054	98.0	98.3
C4	L3 & Traditional	All	−0.61	−1.47***	0.86**	0.39	0.110	98.1	98.4
			Panel D: Volatility portfolios						
U	None		2.03**	−4.15***	6.17***	1.93	—	93.5	94.0
C1	Traditional	Market	1.35*	−2.20	3.55*	1.81	0.003	94.0	94.8
C2	Traditional	All	−0.15	−1.73	1.58	1.39	0.001	97.1	96.5
C3	L3	All	−0.08	−1.69*	1.61	1.11	0.003	97.8	97.7
C4	L3 & Traditional	All	−0.33	−1.84**	1.51	1.02	0.004	98.0	98.0

(*continued*)

Table V—*Continued*

Case	Instruments	Loading(s)	Parameter estimates α_L^C	α_S^C	α_Δ^C	$\sigma(\alpha_\Delta^C)$	$p(\alpha_\Delta^C = \alpha_\Delta^U)$	Model fit R_L^2	R_S^2
			Panel E: Industry concentration portfolios						
U	None		0.13	−1.24***	1.37**	0.63	—	96.2	99.3
C1	Traditional	Market	0.19	−1.22***	1.41**	0.62	0.700	96.2	99.3
C2	Traditional	All	−0.13	−1.35***	1.22**	0.61	0.286	96.6	99.4
C3	L3	All	−0.50	−1.26***	0.76	0.52	0.042	97.2	99.4
C4	L3 & Traditional	All	−0.50	−1.32***	0.82	0.53	0.080	97.3	99.4
			Panel F: Return gap portfolios						
U	None		−0.10	−1.86***	1.76***	0.50	—	96.4	97.1
C1	Traditional	Market	0.14	−1.75***	1.90***	0.51	0.860	96.5	97.1
C2	Traditional	All	−0.28	−1.89***	1.61***	0.51	0.293	97.0	97.4
C3	L3	All	−0.71	−2.13***	1.42***	0.39	0.129	97.4	97.8
C4	L3 & Traditional	All	−0.64	−2.12***	1.48***	0.39	0.205	97.7	98.0

Table VI: Unconditional alpha decompositions.
The table provides decompositions of unconditional Carhart (1997) four-factor alpha estimates into security selection effects, direct factor timing effects, and volatility timing effects for decile portfolios sorted on lagged R^2, active share (AS), active weight (AW), volatility (Vol), industry concentration (ICI), and return gap ($RetGap$). For each set of portfolios, the conditional alphas and factor loadings are from the corresponding conditional models in Table III. We present the unconditional alpha (α_{Δ}^{U}) for each hypothetical long-short strategy in Panel A. In Panel B, each unconditional alpha is decomposed into a security selection component (α_{Δ}^{C}), four factor timing components, and a total volatility timing component. For the market factor ($R_{MKT,t}$), the factor timing component is estimated as $\text{cov}(\hat{\beta}_{\Delta,t}^{C}, R_{MKT,t})$, where $\hat{\beta}_{\Delta,t}^{C}$ is the portfolio's fitted conditional market loading (e.g., for the R^2 strategy, $\hat{\beta}_{\Delta,t}^{C} = \hat{\lambda}_{L,0} + \hat{\lambda}_{L,1}' Z_{L,t-1}^{MKT} - \hat{\lambda}_{H,0} - \hat{\lambda}_{H,1}' Z_{H,t-1}^{MKT}$). The volatility timing effect for the market factor is $(\bar{\beta}_{\Delta,t}^{C} - \hat{\beta}_{\Delta}^{U})\bar{R}_{MKT,t}$, where $\bar{\beta}_{\Delta,t}^{C}$ is the average conditional loading for the market factor, $\hat{\beta}_{\Delta}^{U}$ is the estimated unconditional market factor loading, and $\bar{R}_{MKT,t}$ is the average return on the market factor. The factor timing and volatility timing effects for the size factor ($R_{SMB,t}$), value factor ($R_{HML,t}$), and momentum factor ($R_{UMD,t}$) are estimated analogously. We evaluate the statistical significance of the timing effects in Panel B using a bootstrap approach, and ***, **, and * indicate that the 99%, 95%, and 90% bootstrap confidence interval, respectively, of the corresponding estimate does not include zero. Panel C reports the mean and the 1st, 50th, and 99th percentiles of total factor timing across long-short cohort portfolios formed on each predictor variable. All figures are reported in percentage per year.

	Low−High R^2	High−Low AS	High−Low AW	Low−High Vol	High−Low ICI	High−Low $RetGap$
Panel A: Unconditional alpha						
Unconditional alpha, α_{Δ}^{U}	1.91**	1.47*	1.40**	6.17***	1.37**	1.76***
Panel B: Decomposition						
Conditional alpha, α_{Δ}^{C}	0.34	0.07	0.67*	1.61	0.76	1.42***
Factor timing:						
$\text{cov}(\beta_{\Delta,t}^{C}, R_{MKT,t})$	0.12	0.05	−0.02	2.06***	−0.07	−0.09
$\text{cov}(s_{\Delta,t}^{C}, R_{SMB,t})$	0.18	−0.21	0.17	0.64	−0.06	−0.03
$\text{cov}(h_{\Delta,t}^{C}, R_{HML,t})$	1.02**	0.80	0.31**	2.06	0.21	0.37***
$\text{cov}(u_{\Delta,t}^{C}, R_{UMD,t})$	0.49	0.43	0.30	−0.29	0.42	0.24
Total factor timing	1.82***	1.08	0.75*	4.47*	0.50	0.48*
Total volatility timing	−0.24	0.32	−0.02	0.10	0.10	−0.15
Panel C: Cohort factor timing						
Mean	0.46	0.55	0.13	0.73	0.28	−0.02
P01	0.13	0.13	−0.31	−1.48	−0.03	−0.47
P50	0.45	0.60	0.12	0.64	0.30	−0.03
P99	0.76	0.77	0.60	3.07	0.50	0.51

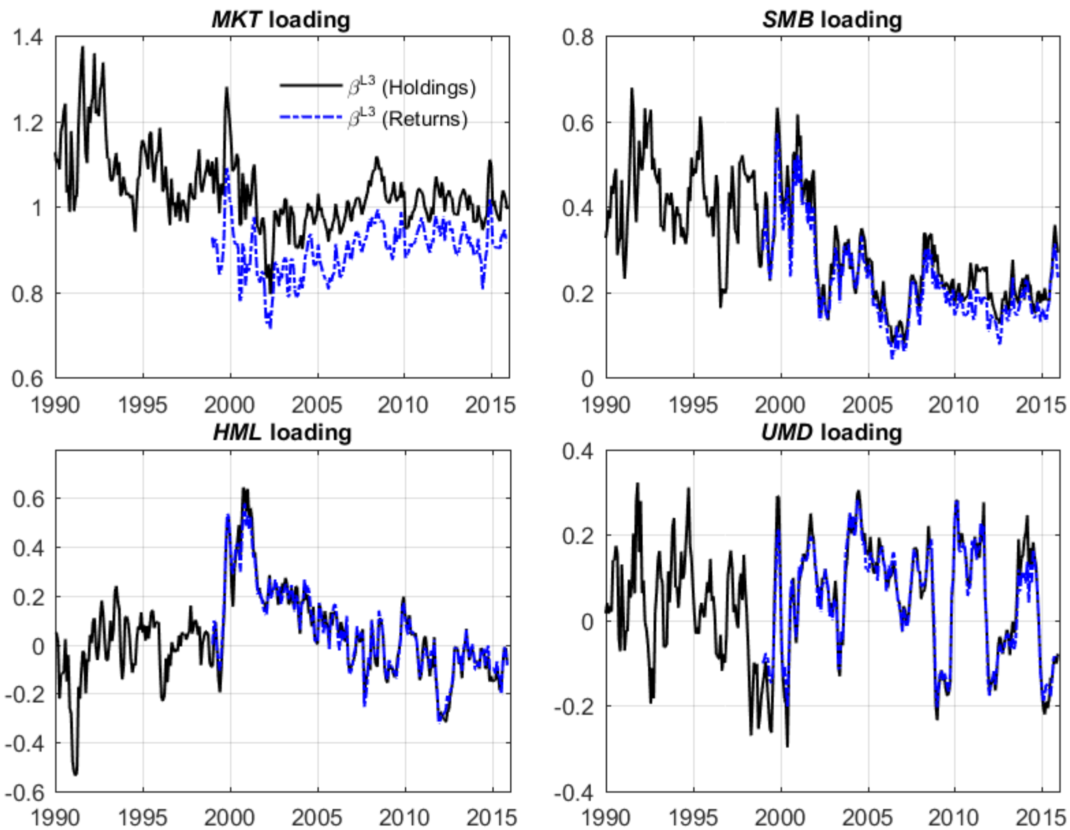

Figure 1: Lagged factor loadings for low-R^2 mutual funds.
The figure shows lagged three-month factor loadings from Carhart (1997) model regressions for the low-R^2 decile portfolio. The solid (dashed) line corresponds to lagged loadings estimated from fund holdings and stock-level excess returns (mutual fund excess returns). In each case, the lagged loadings are the equal-weight lagged loadings across constituent mutual funds. The sample period is January 1990-December 2015.

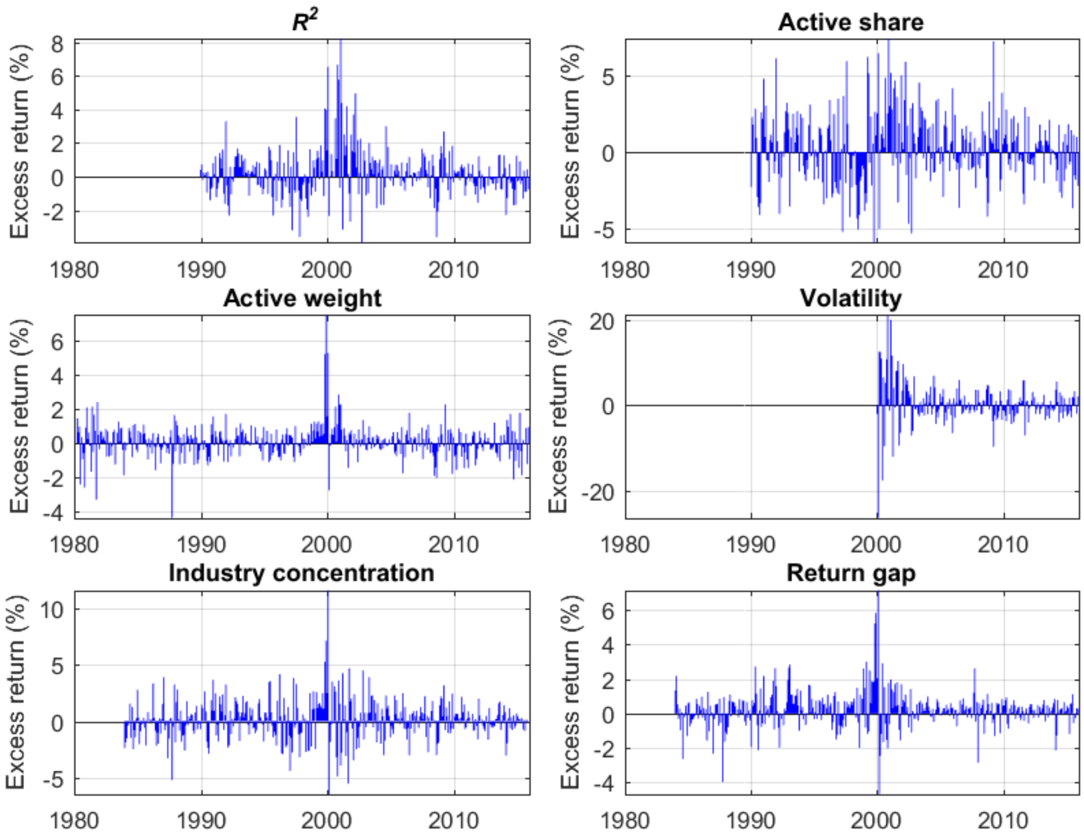

Figure 2: Differences in returns for characteristic-sorted portfolios of mutual funds. The figure shows differences in net returns in percentage per month for the extreme decile portfolios sorted on lagged R^2, active share, active weight, volatility, industry concentration, and return gap.

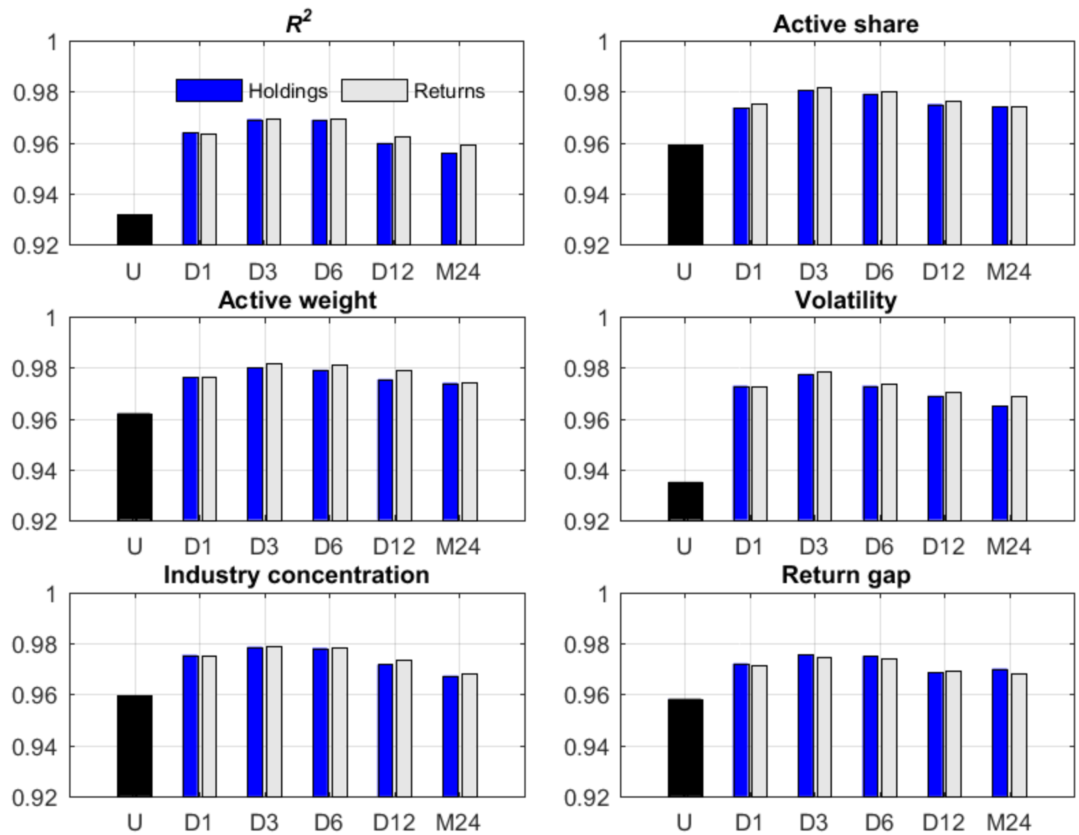

Figure 3: Comparison of model fit.
The figure shows adjusted R^2 values from conditional Carhart (1997) four-factor model regressions for decile portfolios with low R^2, high active share, high active weight, low volatility, high industry concentration, and high return gap. The instruments for a given portfolio are the one-month (D1), three-month (D3), six-month (D6), 12-month (D12), and 24-month (M24) lagged loadings for the corresponding factors. The 24-month instruments are estimated from monthly four-factor regressions, and all other instruments are based on daily four-factor regressions. For each plot, the bar marked "U" corresponds to an unconditional model with no instruments for the factor loadings. For the other cases, the bar on the left (right) corresponds to a conditional model with lagged loading instruments estimated from fund holdings and stock-level excess returns (mutual fund excess returns). The sample period for the volatility strategy is January 2000-December 2015, and the sample period for the remaining strategies is January 1999-December 2015.

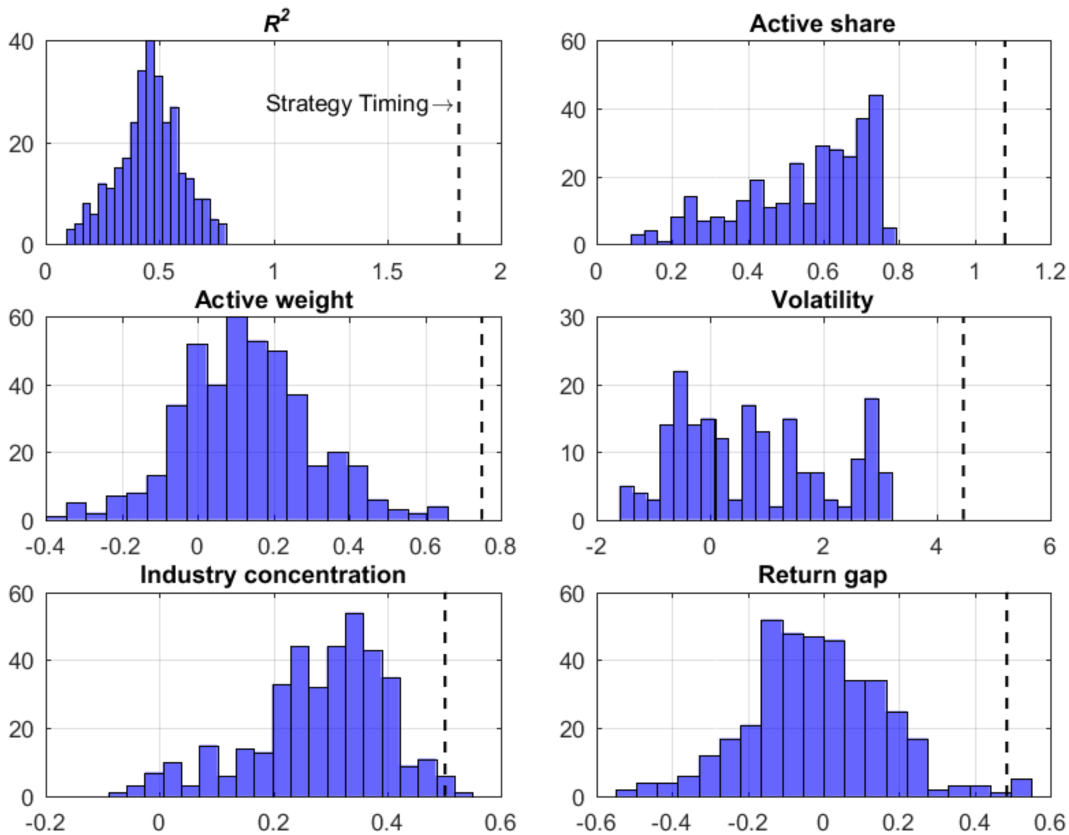

Figure 4: Factor timing for cohort portfolios.
The figure shows histograms of total factor timing for long-short cohort decile portfolios sorted on lagged R^2, active share, active weight, volatility, industry concentration, and return gap. Cohort portfolios are investments in funds with extreme values of a given predictor in the specified month. For example, the January-1990 cohort for the R^2 strategy is a full-sample trading strategy that takes a long (short) position in the decile of mutual funds with the lowest (highest) R^2 values at the beginning of January 1990. The dashed line in each plot shows total factor timing for the overall long-short strategy, which is rebalanced at the beginning of each month. The units on the horizontal axis are in percentage per year.

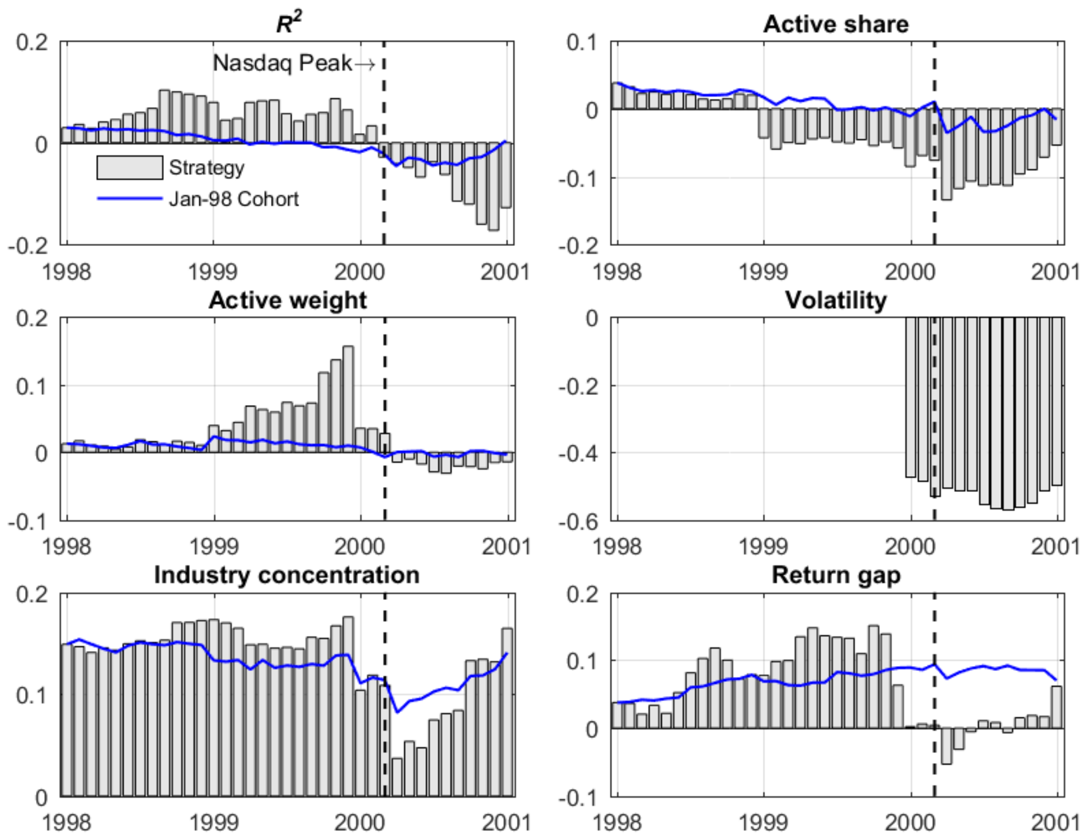

Figure 5: Differences in holdings of technology stocks.
The bars in each plot show the differences in the average weight in technology stocks for long and short decile portfolios formed on the indicated predictor variable. The solid line corresponds to this difference for the January-1998 cohort. Following Brunnermeier and Nagel (2004), technology stocks are defined as Nasdaq firms with price-to-sales ratios in the top quintile.