An Examination of Mutual Fund Timing Using Monthly Holdings Data

Edwin J. Elton*, Martin J. Gruber*, and Christopher R. Blake*

February 2009

* Nomura Professors of Finance, Stern School of Business, New York University
** Joseph Keating, S.J., Distinguished Professor, Fordham University

ABSTRACT

In this paper we use data on the monthly holdings for a set of mutual funds to study the timing ability of these funds. By examining monthly holdings we are able to see how management changes the risk parameters and industry holdings in a fund and to examine how this contributes to timing. We find evidence that timing decisions result in a decrease in performance, whether timing is measured using conditional or unconditional sensitivities. Likewise, sector rotation decisions also result in lower returns. Examining the results for individual sectors shows that the majority of the negative impact on returns from sector rotation comes about because of a fund changing exposure to high-tech stocks.

JEL Classification G11, G12
Keywords: mutual funds, portfolios, composition, timing

INTRODUCTION

While a large body of literature exists on whether active portfolio managers add value, the vast majority of this literature has concentrated on stock selection.[1] In its simplest terms, this literature examines how much better a manager does than holding a passive portfolio of securities with the same risk characteristics (sensitivities to one or more indexes). The bulk of the literature on performance measurement ignores whether managers can time the market as a whole or time across subsets of the market. By doing so, that literature assumes that either timing does not exist or, if it does exist, that it will not distort the measurement of an analyst's ability to contribute to performance through stock selection.

A number of articles have shown that the existence of timing on the part of management can lead to incorrect inference about performance based on either single-index or multiple-index tests of performance.[2] Because of this possibility, and because of the importance of timing ability as an issue, some papers have been written that explore the ability of managers to successfully time the market. This literature started with the work of Treynor and Mazuy (1966), who explored whether there was a non-linear relationship between the beta with the market and the return on the market. This work was followed by Henriksson and Merton (1981), who looked at changes in betas as a reaction to discrete changes in the market return relative to the Treasury bill rate. Other studies followed, using more sophisticated measures of the return-generating process,

[1] See, for example: Elton, Gruber and Blake (1996), Gruber (1996), Daniel, Grinblatt, Titman and Wermers (1997), Carhart (1997) and Zheng (1999), and references therein.
[2] See, for example, Dypvig and Ross (1985) and Elton, Gruber, Brown and Goetzman (2007) for discussion on how timing can lead to incorrect conclusions about management performance.

to examine how time-series sensitivities of mutual fund returns vary with market and factor returns.[3]

The potential problem with almost all of these studies is that they assume management implements timing in a specific way. (E.g., Henriksson and Merton (1981) assume a different but constant beta, according to whether the market return is lower or higher than the risk-free rate). If management chooses to time in a more complex manner, these measures may not detect it. To overcome the estimation problem caused by the assumption of a specific form of timing, we collect data on the actual holdings of mutual funds at monthly intervals. This allows us to construct the beta or betas on a portfolio at the beginning of any month using fund holdings. As explained in more detail later, this is done by using three years of weekly data to estimate the betas on each stock in a portfolio and then using the actual percentage invested in each security to come up with a portfolio beta at a point in time. We refer to portfolio betas constructed this way as "bottom-up" betas.

The advantage of using holdings data at monthly intervals is that we can examine at monthly intervals exactly what decisions management is making. We can observe shifts in the physical securities held in the portfolio (e.g., stocks, bonds, etc.), shifts in the proportions invested in various industries, as well as shifts in the sensitivity to one or more indexes over time.

This approach differs from that which has been taken in the literature with respect to timing measures with the exception of one article: Jiang, Yao and Yu (2007). While our paper follows in the spirit of their article, it differs in several ways. First, Jiang, Yao and Yu only investigate the effect of changing betas in a single-index model. In addition to the one-index

[3] See, for example: Bollen and Busse (2001), Chance and Hemler (2001), Comer (2006), and Ferson and Schadt (1996).

model, we examine a two-index model which recognizes bonds as a separate vehicle for timing, the Fama-French model, and a model that examines the impact of changing allocation across industries.[4] As we show, the use of a more complete model leads to conclusions that are opposite to those reached when the single-index model is used. The reasons for these differences are explored in this paper. Second, we will examine monthly data rather than quarterly holdings data as Jiang, Yao and Yu did. The use of quarterly data misses 18% of the round-trip trades made by the average fund manager.[5] Third, perhaps most importantly, we will account for timing using a full set of holdings including bonds, non-traded equity, preferred stock, other mutual funds, options, and futures. The database used by Jiang, Yao and Yu forced them to assume that all securities except traded equity have the same impact on timing. In particular, they assume the beta on the market of all securities that are not traded equity is zero. Thus non-traded equity, futures, options, preferred stock and mutual funds are all treated as identical instruments, each having a beta on the market of zero.

In the first part of this paper we examine the ability of monthly holdings data to detect timing ability using unconditional betas. We follow this with a section that examines the impact of measures of timing ability that are conditional on publicly available data. Following the general methodology of Ferson and Schadt (F&S), we find that employing a set of variables that measures public information explains a large part of the action management takes with respect to systematic risk and changes the conclusions about timing ability. This is direct evidence that mutual fund management reacts to macro variables that have been shown to predict return and

[4] We report results for the two-index model. The results, while similar to the results for the one-index model do vary for certain funds that hold bonds. We also examined the Fama-French model with the Carhart momentum factor added. The conclusions reached are similar to the ones reported without the momentum factor.

[5] See Elton, Gruber, Krasny and Ozelge (2006) for details on the amount of trades missed using different frequencies of holding data.

also provides additional evidence that using holdings data to measure management behavior is appropriate.

This paper is divided into six sections. The first section discusses out sample. This is followed by a section discussing the methodology we used. Those sections are followed by three empirical sections presenting, respectively, timing results using unconditional betas, industry exposure, and conditional betas. The final section presents our conclusions.

I. Sample

Data on the monthly holdings of individual mutual funds were obtained from Morningstar. Morningstar supplied us with all of its holdings data for all of the domestic (U.S.) stock mutual funds which it followed during the period 1994 to 2004. The only holding Morningstar does not report is that of any security that represents less than 0.006% of a portfolio and, in early years in our sample, holdings beyond the largest 199 holdings in any portfolio. This had virtually no effect on our sample since the sum of the weights almost always equaled one and, in the few cases where it was less than one, the differences were minute.[6]

Previous studies of holdings data have used the Thomson database as the source of holdings data. The Morningstar holdings data are much more complete. Unlike Thomson data, Morningstar data include not only holdings of traded equity, but also holdings of bonds, options, futures, preferred stock, other mutual funds, non-traded equity and cash. Studies of mutual fund behavior from the Thomson database ignore changes across asset categories such as the bond/stock mix and imply that the only risk parameters that matter are those estimated from traded equity securities. While this can affect any study of performance, the drawback of these missing securities is clearly very severe when measuring timing.

---

[6] While Morningstar in early years reported only the largest 199 holdings in a fund, this did not affect our results since most of the funds that held more than 199 securities were index funds and we eliminated index funds from our sample since they do not attempt timing.

Like other studies, the funds in our sample have a high average concentration (over 90%) in common equity. This is used by others to justify using a database that has no information on assets other than traded equity. However, average figures hide the large differences across funds and over time. Twenty-five of the funds in our sample use futures and options, with the future positions being as much as 40% of total assets. Over 20% of the funds vary the proportion put into equity by more than 20%, and they differ in the investments other than equity that are used when equity is changed. The funds that have variation in the percent in equity over time or use assets that can substantially affect sensitivities are precisely the ones that are likely to be timing. Thus, in a study examining timing it is important to have information on all assets the fund holds.

From the Morningstar data we selected all domestic funds, except index and specialty funds, that reported holdings for at least eight months in any calendar year, did not miss two or more consecutive months, and existed for at least two years. These are funds that are reporting monthly holdings most of the time while occasionally missing a month. Only 4.6% of the fund months in our sample did not have data, on average 57% of the years had complete monthly data, and 96% of the fund years were not missing more than two months. Less than 1% of the funds had only eight months of monthly data in any one year.[7] Our sample size is 318 funds and 18,903 fund months.

An important issue is whether restricting our sample to funds with monthly holdings data, which we did throughout this paper, or requiring at least two years of monthly data, introduces a bias. This is examined in some detail in Elton, Gruber and Blake (2008), but a summary is useful here.

---

[7] The data included monthly holdings data for only a very small number of funds before 1998, so we started our sample in that year. In 1998 2.5% of the common stock funds reporting holdings to Morningstar reported these holdings for every month in that year. By 2004 the percentage had grown to 18%.

There are two possible sources of bias. First, funds that voluntarily provide monthly holdings data may be different from those that do not. Second, even if funds that provide monthly holdings are no different from those that do not, requiring at least two consecutive years of holdings data may bias the results. When we require two years of monthly holdings data we are excluding funds that merged and excluding funds that reported monthly holdings data in one year but did not report monthly data in the subsequent year. Each of these potential sources of bias will now be examined.

The first question is whether the characteristics of funds that voluntarily report holdings monthly are different from the general population. Ge and Zheng (2006) examine whether funds that report voluntarily on a quarterly basis are different from those that report semi-annually as required by law in the period they studied. They found that those that reported voluntarily had 0.04% lower expenses, 10% less turnover, were less likely to commit fraud, and differed somewhat in performance. For our study it is the possibility of difference in performance and merger activity that needs to be examined. For each fund in our sample we randomly selected funds with the same investment objective that did not report monthly holdings data. Using the Fama-French model, the difference in average alpha between our sample and the matching sample was three basis points, which is not statistically significant at any meaningful level. We also checked merger activity. There were slightly fewer mergers in the funds that do not report monthly, but in any economic or statistical sense there was no difference.

Another bias could arise by requiring two years of monthly data if funds stopped reporting monthly holdings data because their performance changed or they realized that they were not performing as well as the funds that continued to report monthly data. Of the 104 funds that had 12 months of data in one year but less than 12 months of data in the subsequent year, only four switched to quarterly reporting. Using standard time series regression (the Fama-French model), we again found that they performed no worse than the funds that continued to report holdings on a monthly basis. Finally, 24 funds met reporting requirements in one year and merged in the second. Typical of funds that merge, these were on average poor-performing funds. Examining our measures over the periods they exist shows timing results very slightly below what we report. Thus our measures are very slightly biased upwards. The evidence suggests that our sample does not differ in any meaningful way from the population of funds.

II. Methodology

There are two ways a manager can affect performance beyond security selection. First, the manager can vary the sensitivity of the portfolio to general factors such as the market or the Fama-French factors. This can be done by switching among securities of the same type but with different sensitivities to the factors, or by changing allocation to different types of securities (e.g., stocks to bonds or preferred stocks). Secondly, the manager can vary the industry exposure, overweighting in industries that are forecasted to outperform others (usually called “sector rotation”). Clearly these are interrelated. For example, managers engaged in sector rotation are likely to affect sensitivity to systematic market factors. However, it is useful to examine these separately and then to examine the joint implications of the two types of results.

A. Timing as Index Exposure

One way management can make timing decisions is to change the sensitivity of the portfolio to a set of aggregate indexes that affect returns. Because we have monthly data, we can measure the sensitivity of a portfolio to any influence in successive months over the time period of interest.

A general model for mutual fund returns can be described by a multi-index model of the form

$$R_{Pt} - R_{Ft} = \alpha_P + \sum_{j=1}^{J} \beta_{Pjt} I_{jt} + \varepsilon_{Pt} \tag{1}$$

where R_{Pt} = the return on mutual fund P in month t

R_{Ft} = the return on the 30-day T bill in month t

I_{jt} = the return on index j in month t (see below)

β_{Pjt} = the sensitivity of fund P to index j in month t

α_P = the risk-adjusted excess return on fund P

ε_{Pt} = the residual return on portfolio P in month t

Normally the model is estimated by running a time series regression of the excess return on a fund against the excess return on a set of indexes over time. However, this method suffers from the fact that if management is trying to engage in timing, the β_{Pjt} will vary over time. With holdings data, we can estimate the value of β_{Pjt} at a point in time by calculating the betas for each security in the portfolio and weighting the security betas by the percentage that security represents of the portfolio at that point in time.[8] The betas estimated in this manner are the unconditional betas. It has been argued that there are macro variables that can predict returns, and since the value of the macro variables is known, management should not be given credit for changes in beta in response to these macro variables. Thus we will also estimate conditional

[8] The betas or individual securities are estimated by running regressions on each security against the appropriate index model using three years of weekly data ending in the month being estimated. There is clearly estimation error in the betas of individual securities. This estimation error tends to cancel out and becomes very small when we move to the portfolio level and examine measures over time. See Elton, Gruber and Blake (2008) for a more detailed discussion and for estimates of its effect. The β_{Pjt} is exactly the same as would be obtained if one estimated it using a time series regression with fund returns if the weights remained unchanged over the estimation period.

betas. The exact method used in this estimation will be presented in the section on timing using conditional betas.

We now turn to the problem of choosing the indexes in equation (1). We will first examine the simplest model used in the literature: the single-index model. However, since a number of funds in our sample have significant investments in bonds, we will also use and emphasize a two-index model containing an index of excess returns over the riskless rate for bonds and the excess return for stocks. The third model we use is the familiar Fama French model with the excess return on a bond index added.[9] In Appendix A we describe the details of estimating the models on different types of securities and the procedure we use for missing data.

How do we measure timing? Our timing measure is exactly parallel to the differential return measure used in measuring security selection ability. We examine the differential return earned by varying beta over time rather than holding a constant beta in each period equal to the overall average beta for the history of the fund.

For any model the timing contribution of any variable j is measured by

$$\sum_{t=1}^{T}\left[\beta_{Pjt}-\beta_{Pjt}^{*}\right]\times I_{jt+1}$$

where β_{Pjt}^{*} is the target beta. When we use unconditional betas, the target beta is the average beta for the portfolio over the entire period for which we measure β_{Pjt}. As explained later, when we use conditional betas, the target beta will be the estimated conditional beta. The deviations then become the difference between each month's estimated bottom-up beta and the target beta where the target beta is the expected value of beta adjusted for macro variables. I_{jt+1} is the excess

---

[9] We also added the Carhart momentum factor to this model. The conclusions are not substantially different, and where interesting are presented in the paper. All indexes except for the bond index were provided by Ken French on a weekly basis. The bond index we use is the Lehman U.S. Government/Credit index.

return or differential return for index j for the month following the period over which the beta is estimated. This intuitive measure of timing simply measures how well a manager did by varying the sensitivity of a fund to any particular index compared to simply keeping the sensitivity at its target level. For any fund this can be easily measured for each index or for the aggregate of indexes used in any of the models we explore.

Our measure is very closely related to the measure utilized by Daniel, Grinblatt, Titman and Wermers (1997). While we examine the current beta relative to the average beta, they use as a measure of differential exposure the difference in beta between the current beta and the beta 12 months ago. Each measure has some advantages. We use the average beta because, if the managers have a target beta, the mean is a good estimate of it, and deviation from a target beta is usually what we mean by timing.

B. Changes in Industries Held

The availability of monthly holding data also allows us to look directly at whether changes in the allocations over time across industries improve performance. The methodology directly follows that described in Part A above, but β_{Pjt} is replaced with X_{Pjt} , the fraction of the portfolio P in industry j at time t. The new measure is:

$$\sum_{t=1}^{T}\left[X_{Pjt} - \overline{X}_{Pjt}\right]\times I_{jt+1} \tag{3}$$

where

1. X_{Pjt} is the fraction of mutual fund P invested in industry j at time t.
2. $\overline{X}_{Pjt}$ is the average amount invested in industry j by fund P,
3. I_{jt+1} is the excess return on industry j at time $t+1$ the month following the reported holdings.

We divided equity holdings of the funds into five industry groups as designed by Ken French and available on his web site mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.[10] Since we are interested in changes in stock allocation between industries, we normalize the industry weights at each point in time to add to one.

III. Evidence of Timing Unconditional Betas

Table 1 shows the average difference between the return earned on the factors using the funds' actual betas at the beginning of each month and the return they would have earned if they had held the sensitivity to the factors at their average value over the time period for which we have data. The average difference across funds is broken down into the average difference due to timing on each of the factors and the aggregate of these influences (called "overall"). Table 1 is computed over the 318 funds in our sample. The results for the one-index model are the same as the first index in the two-index model. This comes about because the bond index and stock market index are virtually uncorrelated. Thus, in the interest of space we only present results for the two-index model. For the two-index model the average difference shows positive timing ability of approximately five basis points per month. Examining the components of overall timing for the two-index model shows that this extra return is almost entirely due to the timing of the stock market factor. Of the 318 funds, 233 showed positive timing ability, which is also statistically significantly different from the number we would expect to find by chance. In addition, as shown in Table 2, 56 funds showed statistically significant positive timing ability at the 5% level, while only 10 exhibited significant negative timing ability.

When we examine the four-index model the results are different. The difference in return due to timing the four indexes is minus 11 basis points per month, which is statistically

[10] Similar results were obtained when we used the 17-industry classification designated by French.

significant across the 318 funds. In addition, 96 of the differentials are positive and 22 are negative. Examining Table 2, we also see that only 7 funds show significant positive timing ability at the 5% level, while 54 show significant negative timing ability. Examining the various factors shows that changing betas on the size factor is the major contributor to the negative timing.

Before leaving this section we should consider the possibility that the statistical significance of our results is overstated because the measures across funds may be correlated. This could occur because of common holdings across funds in the same family (see Elton, Gruber and Green (2007)) or because funds in the same family follow common aggregate strategies. To examine the likelihood that the values could have arisen by chance, we performed the bootstrap procedure described in Appendix B. An idea of the significance can be seen by examining Figure 1, which shows the distribution of *t* values for the overall timing measure for the two-index model for the actual 318 funds and for 1,000 random draws for each fund. Note that the *t* values for the 318 actual funds are highly right-skewed. This suggests that there are too many funds showing timing ability to have arisen by chance. A formal test confirms what is visually suggested in Figure 1. The bootstrap procedure shows that the number of funds that show significant positive timing ability at the 5% level should occur by chance less than 0.5% of the time.

Using the four-factor model, Figure 2 presents the *t* values for overall timing measure for that model when we use the 318 actual funds and the 1,000 random funds for each fund. The distribution appears significantly left-skewed. When we formally test significance, the visual picture is confirmed. The number of funds that show significant negative timing ability at the 5% level would occur less than 0.4% of the time. Similar results occur for other cutoffs.

The results from the two- and four-index model are completely different. The timing measure results combined with the simulation indicates that, if one uses a one- or two-index model, mutual funds on average appear to exhibit positive timing ability at an economic and statistical level. When the Fama-French model is used, there is no evidence of successful timing ability on the part of mutual funds.

IV. Differences in Estimates of Market Timing

In this section we will present evidence on why the four-index model is a more appropriate measure of market timing than the two-index model. Let us start by examining two extreme ways management might be attempting to make timing decisions.

In the simplest approach, managers might be only making timing decisions on the sensitivity (beta) of the portfolio with the market and inadvertently neglecting the impact of their decisions on the other common factors that affect return, such as the change in the value/growth characteristics of the portfolio. Whether or not we believe these are equilibrium factors, there is ample evidence that there are differential returns on value and growth and small and large firms that affect fund returns. Thus, inadvertently or not, changing sensitivities to these factors affects fund returns. Furthermore, as we show below, the market sensitivity of a portfolio is highly correlated with sensitivity to one of the other indexes that affects return. Without management action to control the sensitivities to other factors, a change in the market beta will change the other factor sensitivity, and only examining the change in market beta will not correctly measure the total impact on return of a change in the market beta.

The other extreme is to assume that management is concerned with the impact on fund returns of changes in the sensitivity to all four indexes in the return-generating process. In this case, the overall four-index timing measure is appropriate because it measures the impact of changes in all the sensitivities in the return-generating process on returns.

In either case, the correct measure of the impact of management timing decisions should be measured by the four-index model, not by the two-index model.[11]

A. Evidence of Interdependency

In this section we will show that management's market timing choice has a direct effect on their estimated timing choice for other factors. While the additional variables in the Fama-French model were designed to minimize the correlation with the market, the high minus low book-to-market factor (value minus growth) still has considerable correlation (0.591) with the market. If we orthogonalize this factor to the market, the measure of the market timing from the first index of the four-index model is the same as the value of market timing from the first index of the two-index model, but the measure of timing associated with the value minus growth choice changes from -0.0261 to -0.0801. This has two interpretations.[12] If management is only paying attention to the market beta, then the inadvertent change in the timing associated with the value minus growth choice is much more harmful than the prior table would indicate. The -0.0801 is the impact of the inadvertent change in the value minus growth choice, while the -0.0261 is the net result of the inadvertent change in the value minus growth sensitivity and the market timing choice that caused this change. If the fund is paying attention to sensitivity on all indexes, then the poor timing choice on the value minus growth factor is partially mitigated by its impact on market timing. However, management on net is making poor timing choices.

[11] The results for the four- and five-index model are similar. We emphasize the four-index model because while funds make decisions to change the growth or size posture to aid in timing, we know of no funds that change momentum as a timing device.

[12] The timing measure on size and bonds also had small changes.

B. Some Direct Evidence

Our sample is likely to be comprised of funds that are attempting to time and those that are not. If the two-index model or four-index model is correct, then we should find that funds that are attempting to time should be identified as timers by these models. If the manager is successful in timing, the timing measure should be positive; if perverse it should be negative.[13] In either case the absolute value of the *t* associated with the timing measure should be larger if a fund engages in larger timing bets. We can use this to examine whether the results from the two-index model are more sensible than those from the four-index model. A sensible measure should find more evidence of timing for groups that are likely to include more managers who are attempting to time.

We use three metrics to try to measure the size of the timing bets management is making. The first is the standard deviation of the percentage of a fund invested in equities. The second is the amount of turnover. Finally, we separate out mutual funds that declare they are making big timing bets (asset allocation funds) and compare their results with the remaining funds in our sample. The appropriate measure of timing should be able to detect (identify) funds that make large bets.

To detect timing bets and compare models of timing we regressed the absolute value of the significance of the timing variable (*t* value) against the standard deviation of the percentage invested in equity. The results are shown in Part A of Table 3. We would expect managers who practice timing to make large bets and to end up in one of the two tails of the distribution for a properly measured timing variable. As shown in Part A of Table 3, funds that have a higher variance in the percent they place in equity tend to end up as either good timers or bad timers

[13] The counter case, which we consider much less likely, is that timing is completely random or benign for all mutual funds and thus the amount attempted by management should have no effect on the timing measure we are about to propose.

(have large positive or negative t's) when timing is measured using the four-index model. When timing is judged by the two-index model, there is no evidence that funds which vary the percent in equity are identified by the model as timers. Similar results occur when management's attempt to time is proxied by turnover (Panel B, Table 3). Whether we proxy management's attempts to time by the standard deviation of the changes in equity proportions or turnover, the four-index model shows evidence of timing for funds that are likely timers while the two-index model does not. The fact that this influence does not show up when we measure timing ability from the two-index model is further evidence that the two-index model is a less appropriate measure of fund timing.

An alternative way to measure a fund's propensity to time is by its objective. If a fund declares it is going to time by listing itself as an asset allocation fund, then we would expect a properly constructed measure of timing to show more evidence of both attempting to time and timing ability. Table 4 shows the average timing measure for asset allocation funds and non-asset allocation funds as well as the absolute value of the t on the timing measure. When we use the four-index model to measure timing we get the results we would expect, asset allocation funds have larger absolute t's, engage in more attempts at timing, and show more timing skill than non-asset allocation funds.[14] Using the two-index model, asset allocation funds show a smaller propensity to time than do other funds, and are less successful in their timing effect. Since we know these funds are engaging in some timing, this is consistent with the two-index model being the less appropriate model for measuring timing.

[14] Asset allocation funds show a higher propensity to time than do other funds. The average standard deviation of the percent these funds have in equity is 0.078, while for the rest of our sample it is 0.030.

V. Industry Timing

As discussed earlier, a manager can add value by correctly estimating factor returns and switching the exposure to the factor in anticipation of the change in the factor return. A manager can also potentially add value by switching exposure to industry categories. The availability of holding data allows us to explore whether managers have the ability to add value through changing their exposure to industry categories.

We accepted as a definition of relevant industries Ken French's five-industry grouping of firms. The advantage of this definition is not only that French provides a rational and clear definition of the indexes, but he also provides a long history of a return series calculated for each industry. Once again we measured the manager's ability to successfully engage in industry timing (sector rotation) as the difference between the actual exposure at the beginning of the month minus the average exposure over the history of the fund times the return on the industry over the month. These monthly differential returns are accumulated over the full history of each fund. Table 5 provides the overall measure of timing ability along with the timing ability with respect to each industry.

As shown in Table 5, the overall timing measure from industry timing is negative and statistically significantly different from zero whether we judge the average value by the mean or the median. The mean is 33% higher than the median, which is caused by the distribution being left-skewed and including some extremely poor timers. The bulk of the poor timing comes from bad decisions on one industry: high tech. When we examine the mean, 64% of the negative overall timing is due to changing investment in high-tech stocks, while if we examine the median, 63% is due to changing investment in high-tech stocks. Management again seems to exhibit negative timing ability, and the bulk of this negative timing ability comes from one industry: high tech.

Earlier we found that timing on the value-growth factor was a major component of the negative overall timing on the Fama-French factors. Maybe this was due in large part to the timing of mutual funds' investment in the high-tech sector. To examine this, we ran a regression of the Fama-French *HML* (value-growth) index returns on the five French industry sector portfolio (*S*) returns. The regression results are:

$$HML = 0.817 + 0.367S_1 + 0.290S_2 - 0.483S_3 + 0.236S_4 - 0.354S_5$$

t (2.46) (1.89) (2.27) (−7.35) (3.27) (−1.37)

with a coefficient of determination of 0.46.

The size and t value of the high-tech portfolio (S_3) sensitivity and the average value of the returns in the high tech industry group suggest that timing decisions by funds in the high-tech industry strongly influenced the timing results from the four-factor model.[15]

To examine more directly the impact of decisions about high-tech stocks on the timing measures using the four-factor model, we reproduced timing measures for our sample of mutual funds excluding all stock in the high tech-industry (industry 3). Weights were recalculated to maintain full investment. The results are presented in Table 6 along with the previous results from Table 1. Note that the overall mistiming measured by the model is reduced by almost 50%, while the mistiming on the value-growth factor changes sign. With high-tech stocks included, management mistimed the value-growth factor at a statistically significant negative level. If these stocks are excluded from the portfolios, management shows statistically significant positive

[15] We also ran regressions of the market index and the size index against the five industry indexes. The market was significantly loaded on all of the industries, and the size index had no statistically significant coefficient with any of the industry indexes.

timing ability with respect to the value growth factor.[16] Thus mistiming of the tech stocks explains a large amount of the overall negative timing shown by the four-factor model and in particular the negative timing of the Fama-French value-growth factor.

This analysis points out the advantages of employing holdings data. Timing performance can be decomposed to a level that allows the structure of timing mistakes (or accomplishments) to be understood. By combining multifactor analysis with industry analysis, the reason that funds appear to be good timers or bad timers can be better understood.

V. Conditional Betas and Timing

Ferson and Schadt (hereafter F&S) have explored the impact on mutual fund performance of conditioning betas on a set of predetermined time-varying variables representing public information. F&S find that that conditioning beta on a small set of variables changes many of the conclusions about the selection and timing ability of mutual fund managers. They study timing in the context of a single-index model where the parameters of the model are measured from a time series regression of fund returns on market returns using both unconditional betas and betas conditioned on a set of variables measuring public information.

In previous sections we have examined the use of monthly bottom-up betas to measure timing. If changes in these bottom-up betas really measure management action over time and F&S are right that management changes its action based on a set of public-information variables, then these bottom-up betas should be strongly related to the F&S variables. We examine these hypotheses in this section. The section can be thought of as a joint test of the efficacy of bottom-up betas as a measure of management behavior and the efficacy of the F&S variables in explaining management behavior.

---

[16] Looking at the betas of the funds on the high-tech industry, it is clear the funds added high-tech stocks late in the boom and were late in getting out. Recall our sample period coincides with the high-tech bubble.

A. The Conditioning Variables

We follow F&S in defining four variables to capture public information that might affect management's choice of beta.[17] The variables are:

1. The one-month Treasury bill yield lagged one month. To measure this we used the 30-day annualized Treasury bill yield from the CRSP risk-free rates file. This yield is the rate on the bill that matures closest to 30 days.

2. The dividend yield of the CRSP value-weighted index of NYSE/AMEX stocks lagged one month. This is derived by dividing the previous 12 months of dividends by the price level of this index.

3. The term spread lagged one month. This is measured by the yield on a constant maturity 10-year Treasury bond minus the yield on a three-month Treasury bill.

4. The quality (credit) spread in the corporate bond market lagged one month. This is measured by the BAA-rated corporate bond yield less the AAA corporate bond yield.

We follow F&S in assuming that time-varying betas in the four-index model are a linear function of the four conditioning variables discussed above. If we designate these conditioning variables as Z_1 to Z_4 then the conditional beta with respect to any beta for fund P is found from the following time series regression:

$$\beta_{Pjt} = C_{P0j} + \sum_{k=1}^{4} C_{Pkj} Z_{kt} + \varepsilon_{Pjt} \qquad (4)$$

[17] F&S also use a January dummy but find that it has virtually no effect, so we don't include it here.

where

β_{Pjt} is the bottom-up beta for portfolio *P* with respect to factor *j* at time *t* (which does not incorporate conditional information),

C_{Pkj} is the regression coefficient of the *j*th factor on conditioning variable *k* for portfolio *P*,

Z_{kt} is the value of conditioning variable Z_k at time *t*,

ε_{Pjt} is the random error term of the bottom-up beta for portfolio *P* with respect to factor *j* at time *t*.

B. The Impact of Conditioning Variables on Management Behavior

In order to examine whether management was changing beta in reaction to public information, we regress the bottom-up betas with respect to each factor against the F&S conditioning variables for each fund. The results are presented in Table 7.

Panel A shows the average (across all funds) coefficient of determination (R^2) for each of the bottom-up betas for the Fama-French factors and the bond beta. For each of the Fama-French betas and the bond beta, between 25% and 56% is explained by the F&S conditioning variables. This is strong direct evidence that the F&S variables matter on average in explaining how funds change their betas. For the 318 funds, the F&S variables significantly reduced the unexplained variance of the bottom-up market beta 296 times, the small-minus-large beta 308 times, and the value-minus-growth beta 307 times, all at the 5% level. Not all funds include bonds in the portfolios. For the 206 funds that include bonds, the bond betas were significantly related to the F&S variables 72% of the time.

Of the four F&S variables, the variable that was most often significant was credit spread. Credit spread was significantly related to the market and small-minus-large betas and the relationship was primarily negative, while for the value-growth beta and the bond beta, the relationship was primarily positive. Thus when the credit spread widened, funds generally lowered their exposure to the market, small firms and growth stocks while increasing their exposure to large stocks, value stocks and bonds.

The second-most important variable was dividend over price. An increase in this variable led funds to increase their small- and growth-stock exposure while lowering the exposure to large and value firms. In general, an increase in T-bill rates led to an increase in exposure to value stocks relative to growth stocks, while an increase in the term premium caused funds to move from value to growth stocks.

Not all of the signs and significance are consistent with empirical evidence of what predicts higher market returns. Thus the F&S variables capture a mixture of funds using public information and research findings to predict market returns and simply behavioral reaction to macro variables. The justification for removing the effect of changing beta using the F&S variables from the time pattern of beta changes is to not give funds credit for the impact of using public information on their actions. Insofar as the variables simply capture funds' reaction to a change in an economic variable and their reaction is inconsistent with what evidence shows predicts returns, this relationship should not be removed from the time series of betas. Thus both the conditional and unconditional timing measures give insight into funds' timing behavior.

C. Timing Using Conditional Beta

The purpose of this section is to examine whether management timing actions, separate from their reaction to public information, add value. Throughout this section we concentrate on the four-index model. We measure timing using equation (2) but define the target beta (β^*_{Pjt}) as

$\beta^*_{Pjt} = \hat{C}_{P0j} + \hat{C}_{Pkj} \times Z_{kt}$, where the hats indicate regression estimates from equation (4).

Table 8 shows that using conditioning information, the size and significance of the overall timing measure, while still negative, is reduced from the unconditional measure shown in Table 1. The overall timing measure, while much closer to zero, still indicates some negative timing ability and the difference from zero is still significant.

In the prior section we showed that the principal reason for the negative timing measure was the funds' attempted timing of the tech bubble. We can repeat the analysis of Table 8 leaving out tech stocks. When we so do, we obtain an overall timing measure that is positive (0.0008), is exceedingly small, and has a t value of 0.14, indicating that it is indistinguishable from zero at any meaningful level.

We find, as did F&S, that when using conditioning variables, the evidence of perverse timing is greatly diminished, and any perverse timing that remains is entirely due to the choices made in tech stocks during the period of the high-tech stock bubble. These results hold using a different methodology to measure timing as well as a different sample and different time period than those used by F&S.

D. Estimates Using the Time Series of Returns

Fearson and Schadt showed that conditional betas explained more of the time series of fund returns than did unconditional betas for a sample of 67 funds. In this section we will examine the same issue. We will then see how much of the variation of bottom-up conditional betas is explained by top-down conditional betas. The average adjusted R^2 from regressing fund returns on the four-factor model was 0.85. When conditional bottom-up betas are used, the adjusted R^2 increases to 0.884. This increase is very similar to the increase found by F&S. The conditional betas decrease unexplained variance by about 23%. Of the 318 funds, the conditional beta increased the explanatory power at a statistically significant level (using a 5% cutoff rate) for 161 funds.

How similar are the conditional beta estimates using bottom-up betas and top-down betas? We examined this in two ways. First we simply regress for each firm the bottom-up betas on the conditional top-down betas. Second, we looked at consistency in sign and significance between the coefficients of the conditional top-down and bottom-up betas. When we regress in time series the bottom-up betas on the conditional betas estimated from the time series of fund returns, we get R^2 on average across all funds ranging from 0.18 for the beta on the market to 0.14 for the beta on the small-minus-large factor. When the bottom-up betas were regressed directly on the F&S variables, the F&S variables captured about 50% of the variation in bottom-up betas. Thus conditional variables estimated from a time series of returns captures about one-third of the variation in bottom-up betas that is explained by using the F&S variables directly. Another way to examine the relationship of the top-down conditional betas and the bottom-up conditional betas is to examine whether they capture the same relationship. 1,724 slope coefficients were significant when bottom-up betas were regressed on the F&S variables: 328 of the coefficients on the F&S variables were significant when the F&S variables were included in the regression of fund returns on the indexes, and 131 were significant in both sets of regressions. Of the 131 that were jointly significant, 83 had the same sign. Thus although the conditional variables capture a substantial amount of the variation in bottom-up betas and top-down betas, the relationships are quite different.

VI. Conclusion

In this paper we use data on the monthly holdings for a set of mutual funds to study the timing ability of these funds. By examining monthly holdings we are able to see how management changes the risk parameters and industry holdings in a fund and to examine how this contributes to timing.

Our study differs from previous studies in both the methodology used and in the accuracy of the data. Other studies that use holdings data have employed a database that includes only data on the holdings of publicly traded stock. Our database contains holdings of funds in options, futures, other mutual funds, preferred stock, bonds, and non-traded equity. Many funds use these additional instruments to time, and ignoring their presence can lead to erroneous conclusions about management timing decisions. Furthermore, the few studies that use holdings data have used quarterly data, which gives at best a much coarser measure of timing.

Our major results are based on the Fama-French model with the addition of a bond index. A portfolio's "bottom-up" betas with respect to any given index at a point in time are calculated by multiplying the betas on each security in a portfolio by the fraction that security represents of the portfolio and then summing across all the securities held by the fund. In addition, we extend the work of F&S to calculate conditional betas based on observable macro variables.

We find evidence that timing decisions result in a decrease in performance, whether timing is measured using conditional or unconditional sensitivities. Likewise, sector rotation decisions also result in lower returns. Examining the results for individual sectors shows that the majority of the negative impact on returns from sector rotation comes about because of a fund changing exposure to high-tech stocks. The funds in our sample invested in high-tech stocks late in the bubble and continued to invest heavily after it broke. Choices made with respect to high-tech stocks were also a major reason for the negative timing results when the four-index model was used. This occurred in large part because of the correlation of the value-growth index with returns on high-tech stocks. When we removed the effect of high-tech stocks from our data, management timing decisions neither help nor hurt.

We also explored timing using a one- and a two-index model. These models showed positive timing. However, choices on market sensitivity also impacted sensitivity choices on other variables that affect return. When these impacts are taken into account, the average timing measure is negative. Furthermore, using the two-index model led to conclusions that showed asset allocation funds were poorer timers than non-asset allocation funds, and funds with high turnover or high variance of the proportion invested in equity or were asset allocation funds were less likely to time. Opposite results consistent with the expected direction were obtained when the four-index model was used.

Appendix A

Bottom-Up Holdings-Based Estimations

Our sample allows us to estimate the mutual fund beta from holdings data as frequently as monthly. To do this at any point in time, we estimate a time series regression (equation (1)) using three years of weekly past return data on each common stock or mutual fund held by the mutual fund under study. There are two problems. First, if less than 36 months of data are available, we use as much data as is available unless it is less than 12 months. If we have less than 12 months of data available, we set the beta for the stock equal to the average beta for all other stocks in the portfolio. On average this had to be done for less than 1.4% of the securities in any portfolio. The second problem involves the estimation of equation (1) for securities other than common stock and mutual funds.

For T-bills and bonds with less than one year to maturity we set all betas to zero. For each of the following categories of investments: long-term bonds, preferred stocks and convertibles, we used an index of that category and obtained estimated betas by running a regression of the category index against the appropriate model. Each bond, convertible or preferred was assumed to have the same beta as the relevant index. Finally, for options and futures we used the same beta as the underlying instrument adjusted for leverage. We used the Black-Scholes formula to estimate the betas for options.

The beta for any fund can be found at a point in time by weighting the beta on each security held in the fund at that time by the percent that the security represents of the fund's portfolio.

Appendix B

In Table 2 we present the number of funds with significant differential returns. The question is if this number could arise by chance. To examine this, we used a bootstrap procedure. For each fund we used the actual betas each month and actual average betas over the full period. Then for each month we randomly drew with replacement factor returns. These returns are the actual returns (all from the same month across factors) that occurred sometime in the past. The differential return in the month was computed using the actual beta, actual average betas, and the random factor returns. These differential returns are accumulated over the time period and significance is computed using the distribution of differential returns. For each simulation we count the number of funds that show significance at various levels. We repeat this 1,000 times.

Table 1

Mean Differences (in percent)

	Mean	**T-Values**	**Median**
Two-Index			
Overall	0.0520	8.02	0.0740
Market	0.0517	8.03	0.0742
Bonds	0.0003	1.23	0.0
Four-Index			
Overall	-0.1073	-8.41	-0.0515
Market	-0.0247	-6.08	-0.0130
Size	-0.0572	-9.49	-0.0221
Value Growth	-0.0261	-2.88	-0.0213
Bonds	0.0006	2.20	0.0

Table 2

Number of Significant t Values

	At 5% Level		**At 1% Level**	
	Positive	**Negative**	**Positive**	**Negative**
Two-Index				
Overall	56	10	6	3
Market	55	12	5	3
Bond	11	14	0	7
Four-Index				
Overall	7	54	0	6
Market	12	53	2	7
Size	5	20	0	1
Value Growth	16	40	3	7
Bond	11	14	0	8

Table 3

Do Funds That Are Actively Changing the Portfolio Show Up as Achieving Positive or Negative Timing?

A. Panel A Active Timing Proxied by the Standard Deviation of Percent in Equity

Timing Judged by Absolute Value of *t* From	**Regressive Coefficient**	***t* value**	**Adjusted R^2**
Two-Index Model	-0.65	-0.62	0.001
Four-Index Model	3.27	2.96	0.027

B. Panel B Active Timing Proxied by Turnover

Timing Judged by Absolute Value of *t* From	**Regressive Coefficient**	***t* value**	**Adjusted R^2**
Two-Index Model	-0.001	-0.80	0.002
Four-Index Model	0.002	3.07	0.029

Table 4

This table shows the absolute value of the *t* values of our timing measure for each of the two groups of funds.

	#	Two-Index Model		Four-Index Model	
		Timing Measure	$\lvert t \rvert$	Timing Measure	$\lvert t \rvert$
Asset Allocation Funds	29	-0.002	1.048	-0.042	1.053
Non-Allocation Funds	289	0.057	1.186	-0.114	0.911

Table 5

Timing by Industry (%)

	Mean	*t*	Medium	Top Quartile	Bottom Quartile
Overall	-0.0742	-13.10	-0.0556	0.0199	-0.2014
1 Consumer	-0.0096	-7.40	-0.0074	0.0171	-0.3832
2 Manufacturing	-0.0091	-4.95	-0.0084	0.0274	-0.0470
3 High Tech	-0.0476	-8.57	-0.0349	0.0552	-0.1671
4 Health	-0.0018	-1.49	-0.0005	0.0238	-0.0292
5 Other	-0.0062	-3.62	-0.0051	0.0199	-0.2014

Table 6

Timing With and Without Tech Stocks
(Four-Index Model)

	Overall	Market	Size	Value Growth	Bond
Including Tech Stock					
Mean	-0.1073	-0.0247	-0.0572	-0.0261	0.0006
t	-8.41	-6.08	-9.49	-2.88	2.20
Median	-0.0515	-0.0136	0.0221	-0.0213	0.0000
Excluding Tech Stock					
Mean	-0.0557	-0.0365	-0.0424	0.0226	-0.0007
t	-5.77	-8.96	-8.37	2.95	-1.87
Median	-0.0274	-0.0288	-0.0157	0.0005	0

Table 7

Panel A **Panel B**

					Times Significantly Related To			
	Number of Funds	**Significant Improvement In Fit**	**Average Adj R^2**		**T-bill**	**Divided Price**	**Term**	**Credit Spread**
b_{14}	318	296	0.42		94	99	69	190
b_{24}	318	308	0.50		86	100	51	265
b_{34}	318	307	0.56		102	137	63	261
b_{44}	206	148	0.25		37	45	32	93

Table 8

Mean Differences (in percent)

	Mean	t-Values	Median
Two-Index			
Overall	-0.0054	-1,76	-0.0066
Market	-0.0052	-1.71	-0.0058
Bonds	-0.0002	-0.83	0.0
Four-Index			
Overall	-0.0287	-4.32	-0.0174
Market	-0.0055	-2.71	-0.0053
Size	-0.0040	1.83	-0.0069
Value Growth	-0.0271	-4.50	-0.0400
Bonds	-0.0001	-0.58	0.0

Figure 1

**Distributions of t Values:
Actual Sample versus Random Draw
2-Factor Model**

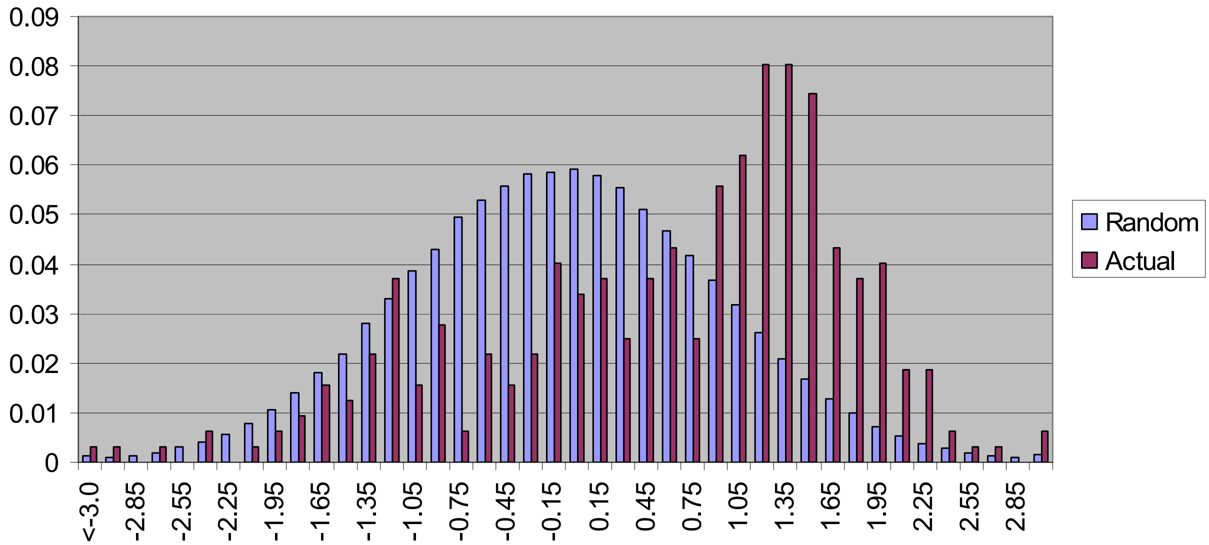
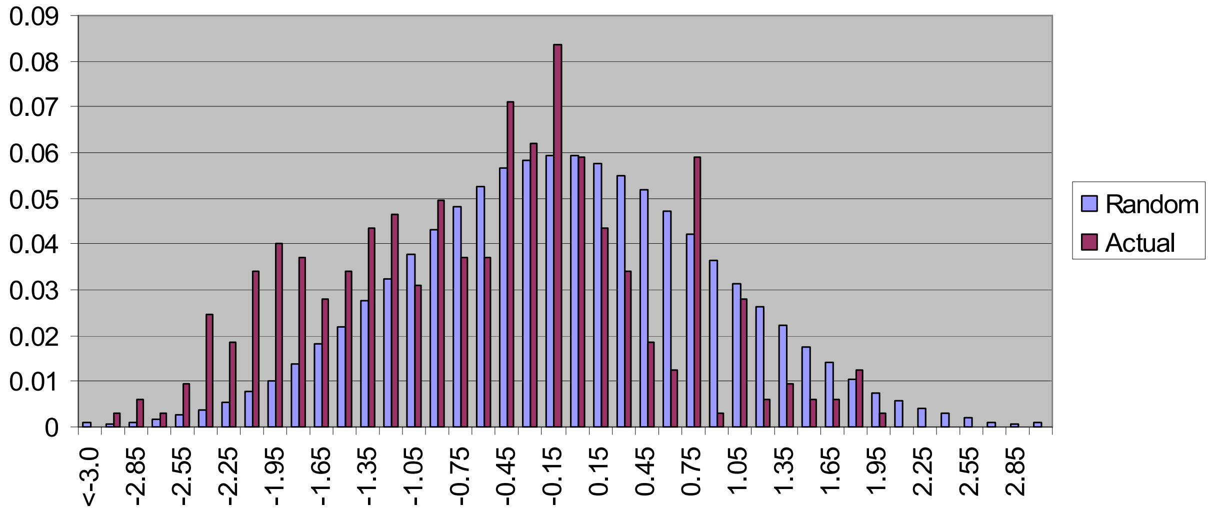


Figure 2

Distributions of t values:
Actual Sample versus Random Draw
4-Factor Model



Bibliography

Becker, Connie; Ferson, Wayne; Mayers, David; and Schill, Michael (1999). "Conditional Market Timing with Benchmark Investors." *Journal of Financial Economics* 52, 119-148.

Blume, Marshall (1971), "On the Assessment of Risk." *Journal of Finance* 26, pp.1-10.

Blume, Marshall (1975), "Betas and Their Regression Tendencies." *Journal of Finance* 30 pp. 785-795.

Bollen, Nicolas P. B. and Busse, Jeffrey A. (2001). "On the Timing Ability of Mutual Fund Managers." *The Journal of Finance* 56: 1075-1094.

Brown, Stephen; Gallagher, David; Steenbeek, Onno; and Swan, Peter (2005). "Double or Nothing: Patterns in Equity Fund Holdings and Transactions." Working Paper, NYU Stern School of Business.

Carhart, Mark (1997). "On Persistence in Mutual Fund Performance." *Journal of Finance,* 52, 57-82.

Chance, Don and Hemler, Michael (2001). "The Performance of Professional Market Timers: Daily Evidence from Executed Strategies." *Journal of Financial Economics* 62, 377-411.

Chen, Yong; Ferson; Wayne; and Peters, Helen (2006). "Measuring the Timing Ability of Fixed Income Mutual Funds. "

Comer, George (2006). "Hybrid Mutual Funds and Market Timing Performance." *Journal of Business* 79(2).

Cumby, Robert and Modist, David (1987). "Testing for Market Timing Ability." *Journal of Financial Economics*, 19: 169-188.

Daniel, Kent; Grinblatt, Mark; Titman, Sheridan; and Wermers, Russ (1997). "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *Journal of Finance*, 52(3), pp. 1035-1058.

Dybvig, Philip and Ross, Stephen (1985). "Differential Information and Performance Measurement Using the Security Market Line." *Journal of Finance* 40: 383-399.

Elton, Edwin J.; Gruber, Martin J.; and Blake, Christopher R. (1996). "The Persistence of Risk-Adjusted Mutual Fund Performance." *Journal of Business* 69 (2): 133-157.

Elton, Edwin J.; Gruber, Martin J.; and Blake, Christopher R. (2008). “Holdings Data, Security Returns and the Selection of Superior Mutual Funds.” Working paper.

Elton, Edwin J.; Gruber, Martin J.; Brown, Steven; and Goetzman, William (2007). *Modern Portfolio Theory and Investment Analysis*. Seventh edition, John Wiley & Sons.

Elton, Edwin J.; Gruber, Martin J.; and Green, Clifford (2007). “The Impact of Mutual Fund Family Membership on Investor Risk.” *Journal of Financial and Quantitative Analysis* 90:6-7 1299-1314.

Elton, Edwin J.; Gruber, Martin J.; Krasny, Yoel; and Ozelge, Sadi (2006). “The Effect of the Frequency of Holding Data on Conclusions about Mutual Fund Management Behavior.” Unpublished manuscript, New York University.

Ferson, Wayne E.; and Schadt, Rudi W. (1996). “Measuring Fund Strategy and Performance in Changing Economic Conditions.” *Journal of Finance* 51: 425-462.

Ge, Weili; Zheng, Lu (2005). “The Frequency of Mutual Fund Portfolio Disclosure.” Working paper, University of California Irvine.

Gruber, Martin J. (1996). “Another Puzzle: The Growth in Actively Managed Mutual Funds.” *Journal of Finance* 51: 783-810.

Henriksson, Roy and Merton, Robert (1981). “On Market Timing and Investment Performance.” *Journal of Business* 54, 513-533.

Jiang, George; Yao, Tong; Yu, Tong (2007). “Do Mutual Funds Time the Market? Evidence From Holdings Data.” *Journal of Financial Economics*, 88:1, 119-145.

Mamayski, Harry; Spiegel, Matthew; and Zhang, Hong (2008). “Estimating the Dynamics of Mutual Fund Alphas and Betas.” *Review of Financial Studies* 21, pp. 233-264.

Myers, Mary Margaret; Poterba, James; Shackelford, Douglas; and Shoven, John (2004). “Copycat Funds: Information Disclosure Regulation and the Returns to Active Management in the Mutual Fund Industry.” *Journal of Financial Economics* 47: 515-541.

Sharpe, William (1992). “Asset Allocation: Management Style and Performance Measurement.” *Journal of Portfolio Management* 18: 7-19.

Treynor, Jack and Mazuy, M. (1966) “Can Mutual Funds Outguess the Market?” *Harvard Business Review*, 44(4), pp. 131-136.

Wermers, Russ (2001). “The Potential Effects of More Frequent Portfolio Disclosure on Mutual Fund Performance.” *Investment Company Institute Perspective*, 1-12.

Zheng, Lu (1999). “Is Money Smart? – A Study of Mutual Fund Investors’ Fund Selection Ability.” *Journal of Finance* 54(3): 901-933.